

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
9 October 2003 (09.10.2003)

PCT

(10) International Publication Number  
WO 03/083039 A2

(51) International Patent Classification <sup>7</sup> :		C12N	US	60/372,141 (CON)
(21) International Application Number:		PCT/US02/21485	Filed on	12 April 2002 (12.04.2002)
(22) International Filing Date:		3 July 2002 (03.07.2002)	US	60/372,967 (CON)
(25) Filing Language:		English	Filed on	16 April 2002 (16.04.2002)
(26) Publication Language:		English	US	60/303,828 (CON)
(30) Priority Data:			Filed on	9 July 2001 (09.07.2001)
60/303,046	5 July 2001 (05.07.2001)	US	US	60/361,748 (CON)
60/303,828	9 July 2001 (09.07.2001)	US	Filed on	5 March 2002 (05.03.2002)
60/304,502	11 July 2001 (11.07.2001)	US	US	60/304,502 (CON)
60/305,011	12 July 2001 (12.07.2001)	US	Filed on	11 July 2001 (11.07.2001)
60/305,262	13 July 2001 (13.07.2001)	US	US	60/324,711 (CON)
60/305,673	16 July 2001 (16.07.2001)	US	Filed on	25 September 2001 (25.09.2001)
60/306,085	17 July 2001 (17.07.2001)	US	US	60/361,178 (CON)
60/307,536	24 July 2001 (24.07.2001)	US	Filed on	1 March 2002 (01.03.2002)
60/308,228	27 July 2001 (27.07.2001)	US	US	60/358,939 (CON)
60/308,877	30 July 2001 (30.07.2001)	US	Filed on	22 February 2002 (22.02.2002)
60/312,203	14 August 2001 (14.08.2001)	US	US	60/322,640 (CON)
60/322,640	17 September 2001 (17.09.2001)	US	Filed on	17 September 2001 (17.09.2001)
60/323,484	19 September 2001 (19.09.2001)	US	US	60/363,683 (CON)
60/323,948	21 September 2001 (21.09.2001)	US	Filed on	12 March 2002 (12.03.2002)
60/323,821	21 September 2001 (21.09.2001)	US	US	60/305,011 (CON)
60/324,711	25 September 2001 (25.09.2001)	US	Filed on	12 July 2001 (12.07.2001)
60/327,893	9 October 2001 (09.10.2001)	US	US	60/305,262 (CON)
60/331,768	21 November 2001 (21.11.2001)	US	Filed on	13 July 2001 (13.07.2001)
60/359,191	21 February 2002 (21.02.2002)	US	US	60/373,051 (CON)
60/358,939	22 February 2002 (22.02.2002)	US	Filed on	16 April 2002 (16.04.2002)
60/360,923	28 February 2002 (28.02.2002)	US	US	60/373,881 (CON)
60/361,178	1 March 2002 (01.03.2002)	US	Filed on	19 April 2002 (19.04.2002)
60/360,830	1 March 2002 (01.03.2002)	US	US	60/305,673 (CON)
60/361,748	5 March 2002 (05.03.2002)	US	Filed on	16 July 2001 (16.07.2001)
60/363,683	12 March 2002 (12.03.2002)	US	US	60/307,536 (CON)
60/363,429	12 March 2002 (12.03.2002)	US	Filed on	24 July 2001 (24.07.2001)
60/372,141	12 April 2002 (12.04.2002)	US	US	60/373,287 (CON)
60/372,967	16 April 2002 (16.04.2002)	US	Filed on	17 April 2002 (17.04.2002)
60/373,051	16 April 2002 (16.04.2002)	US	US	60/323,948 (CON)
60/373,063	16 April 2002 (16.04.2002)	US	Filed on	21 September 2001 (21.09.2001)
60/373,287	17 April 2002 (17.04.2002)	US	US	60/359,191 (CON)
60/373,280	17 April 2002 (17.04.2002)	US	Filed on	21 February 2002 (21.02.2002)
60/373,881	19 April 2002 (19.04.2002)	US	US	60/306,085 (CON)
10/187,975	2 July 2002 (02.07.2002)	US	Filed on	17 July 2001 (17.07.2001)
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:			US	60/323,484 (CON)
US	60/303,046 (CON)	US	Filed on	19 September 2001 (19.09.2001)
Filed on	5 July 2001 (05.07.2001)	Filed on	US	60/323,821 (CON)
			US	60/373,280 (CON)
			Filed on	17 April 2002 (17.04.2002)

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: The present invention relates to novel polypeptides, and the nucleic acids encoding them, having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.



WO 03/083039 A2



US 60/308,228 (CON)  
 Filed on 27 July 2001 (27.07.2001)  
 US 60/308,877 (CON)  
 Filed on 30 July 2001 (30.07.2001)  
 US 60/312,203 (CON)  
 Filed on 14 August 2001 (14.08.2001)  
 US 60/373,063 (CON)  
 Filed on 16 April 2002 (16.04.2002)  
 US 60/360,923 (CON)  
 Filed on 28 February 2002 (28.02.2002)  
 US 60/360,830 (CON)  
 Filed on 1 March 2002 (01.03.2002)  
 US 60/327,893 (CON)  
 Filed on 9 October 2001 (09.10.2001)  
 US 60/331,768 (CON)  
 Filed on 21 November 2001 (21.11.2001)  
 US Not furnished (CON)  
 Filed on 2 July 2002 (02.07.2002)

(71) **Applicant (for all designated States except US): CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants (for US only): LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apt. 1C, Branford, CT 06405 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US). **CATTERTON, Elina** [FI/US]; 584, Boston Post Road, Madison, CT 06443 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Drive, New Haven, CT 06515 (US). **HJALT, Tord** [SE/US]; 514 Main Street, Apartment 30, East Haven, CT 06512 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost

Drive, Branford, CT 06405 (US). **BAUMGARTNER, Jason, C.** [US/US]; 1697 Quinnipiac Avenue, New Haven, CT 06513 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **SMITHSON, Glennnda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **PENA, Carol, E., A.** [US/US]; 604 Orange Street, Number 2, New Haven, CT 06511 (US). **CHAPOVAL, Andrei** [US/US]; 1 Maltby Street, Branford, CT 06405 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US).

(74) **Agent: ELRIFI, Ivor, R.;** Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C, One Financial Center, Boston, MA 02111 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

## THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

### FIELD OF THE INVENTION

The present invention relates to novel polypeptides, and the nucleic acids encoding them, having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

## BACKGROUND OF THE INVENTION

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins, and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected



of suffering from a condition brought on by altered or mis-regulated levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

Antibodies are multichain proteins that bind specifically to a given antigen, and bind poorly, or not at all, to substances deemed not to be cognate antigens. Antibodies are comprised of two short chains termed light chains and two long chains termed heavy chains. These chains are constituted of immunoglobulin domains, of which generally there are two classes: one variable domain per chain, one constant domain in light chains, and three or more constant domains in heavy chains. The antigen-specific portion of the immunoglobulin molecules resides in the variable domains; the variable domains of one light chain and one heavy chain associate with each other to generate the antigen-binding moiety. Antibodies that bind immunospecifically to a cognate or target antigen bind with high affinities. Accordingly, they are useful in assaying specifically for the presence of the antigen in a sample. In addition, they have the potential of inactivating the activity of the antigen.

Therefore there is a need to assay for the level of a protein effector of interest in a biological sample from such a subject, and to compare this level with that characteristic of a nonpathological condition. In particular, there is a need for such an assay based on the use of an antibody that binds immunospecifically to the antigen. There further is a need to inhibit the activity of the protein effector in cases where a pathological condition arises from elevated or excessive levels of the effector based on the use of an antibody that binds immunospecifically to the effector. Thus, there is a need for the antibody as a product of manufacture. There further is a need for a method of treatment of a pathological condition brought on by an elevated or excessive level of the protein effector of interest based on administering the antibody to the subject.

## SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 61. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of

SEQ ID NO:2n, wherein n is an integer between 1 and 61 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

5 In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 wherein said therapeutic is the polypeptide selected from this group.

10 In another embodiment, the invention comprises a method for determining the presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of  
15 polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 in a first mammalian subject, the method  
20 involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample  
25 indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, the method including  
30 introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer

between 1 and 61, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal. The promoter may or may not be the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, the method including introducing a cell sample expressing the polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, the method including administering the polypeptide to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61; a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61; a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and the complement of any of the nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61 that encodes a variant polypeptide,

wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 61.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61; a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61; and a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61, or a complement of the nucleotide sequence.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ

ID NO:2n, wherein n is an integer between 1 and 61, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the  
5 nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid  
10 sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

In another embodiment, the invention involves a method for determining the  
15 presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61 in a sample, the method including providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or  
20 amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

In another embodiment, the invention involves a method for determining the  
25 presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 61 in a first mammalian  
subject, the method including measuring the amount of the nucleic acid in a sample from  
30 the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

The invention further provides an antibody that binds immunospecifically to a NOVX polypeptide. The NOVX antibody may be monoclonal, humanized, or a fully human antibody. Preferably, the antibody has a dissociation constant for the binding of the NOVX polypeptide to the antibody less than  $1 \times 10^{-9}$  M. More preferably, the NOVX antibody neutralizes the activity of the NOVX polypeptide.

In a further aspect, the invention provides for the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, associated with a NOVX polypeptide. Preferably the therapeutic is a NOVX antibody.

In yet a further aspect, the invention provides a method of treating or preventing a NOVX-associated disorder, a method of treating a pathological state in a mammal, and a method of treating or preventing a pathology associated with a polypeptide by administering a NOVX antibody to a subject in an amount sufficient to treat or prevent the disorder.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel



sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

**TABLE A. Sequences and Corresponding SEQ ID Numbers**

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
1a	CG103191-02	1	2	chromogranin A-like
1b	CG103191-03	3	4	chromogranin A-like
1c	CG103191-04	5	6	chromogranin A-like
1d	251425133	7	8	chromogranin A-like
1e	251425611	9	10	chromogranin A-like
1f	278460276	11	12	chromogranin A-like
1g	278456175	13	14	chromogranin A-like
2a	CG105757-01	15	16	Kelch and BTB/POZ containing membrane protein like
3a	CG108175-01	17	18	neurexin III-alpha membrane-bound type 1 precursor like
3b	CG108175-02	19	20	neurexin III-alpha membrane-bound type 1 precursor like
3c	CG108175-03	21	22	neurexin III-alpha membrane-bound type 1 precursor like
3d	CG108175-04	23	24	neurexin III-alpha membrane-bound type 1 precursor like
3e	CG108175-05	25	26	neurexin III-alpha membrane-bound type 1 precursor like
4a	CG108624-01	27	28	protocadherin 68-like
5a	CG108771-01	29	30	Type Ib membrane protein like
6a	CG108782-01	31	32	Transmembrane like
6b	CG108782-02	33	34	Transmembrane like
7a	CG108801-01	35	36	EGF-domain Transmembrane Protein like
7b	CG108801-02	37	38	EGF-domain Transmembrane Protein like
8a	CG109717-01	39	40	Single Pass Transmembrane-Like
9a	CG110477-01	41	42	Desmoglein 3 variant like
10a	CG110540-01	43	44	Pheromone Receptor like
10b	CG110578-02	45	46	Neuralin 2 like
11a	CG110725-01	47	48	Osteopontin like
11b	209934449	119	120	osteopontin-like
12a	CG111683-01	49	50	surfactant protein-C like
12b	CG111683-02	51	52	surfactant protein-C like
12c	CG111683-03	53	54	surfactant protein-C like
13a	CG112655-01	55	56	germ cell-less 1 protein like
14a	CG112813-01	57	58	NK receptor-like
14b	CG112813-02	59	60	NK receptor-like
14c	CG112813-04	61	62	NK receptor-like
14d	CG112813-05	63	64	NK receptor-like
14e	CG112813-06	65	66	NK receptor-like
14f	209886463	67	68	NK receptor-like
14g	277731421	69	70	NK receptor-like

15a	CG112869-01	71	72	Pecanex like
16a	CG113377-01	73	74	G1-related zinc finger protein like
17a	CG113730-01	75	76	nodal precursor like
17b	210982580	77	78	nodal precursor like
17c	CG113794-02	79	80	PA domain containing protein like
18a	CG115187-01	81	82	transmembrane protein like
18b	CG115187-02	83	84	transmembrane protein like
18c	CG115187-03	85	86	transmembrane protein like
18d	262770580	87	88	transmembrane protein like
18e	257788219	121	122	transmembrane-protein like
19a	CG115540-01	89	90	Membrane Protein containing Collagen triple helix repeat like
20a	CG118689-01	91	92	Uroplakin 1b-like
20b	CG118689-02	93	94	Uroplakin 1b-like
21a	CG120748-01	95	96	LMBR1 Long Form like
22a	CG121519-01	97	98	LDL Receptor Domain Containing Protein
23a	CG122176-01	99	100	Fibronectin domain containing protein like
24a	CG122691-01	101	102	Fn3/TSPN/Collagen/vWF domain cotaining protein like
25a	CG122863-01	103	104	Membrane Protein like
25b	CG122863-02	105	106	neurotrimin like
26a	CG50880-04	107	108	Estrogen regulated protein like
27a	CG51812-03	109	110	protocadherin like
28a	CG51923-01	111	112	protocadherin like
28b	CG51923-03	113	114	Protocadherin FAT-like
28c	207756525	115	116	protocadherin like
28d	207756686	117	118	protocadherin like

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to: *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, cellular regeneration, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies,

graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders including autoimmune disorders, hematopoietic disorders, and the various dyslipidemias, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation and fertility.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

**NOVX clones**

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy.

Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an

integer between 1 and 61 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 61; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 61 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61 is changed from

that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

### **NOVX Nucleic Acids and Polypeptides**

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host cell) in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or

phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be

cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term “oligonucleotide” refers to a series of linked nucleotide  
5 residues. A short oligonucleotide sequence may be based on, or designed from, a genomic  
or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical,  
similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides  
comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably  
about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide  
10 comprising a nucleic acid molecule less than 100 nt in length would further comprise at  
least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and  
61, or a complement thereof. Oligonucleotides may be chemically synthesized and may  
also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention  
15 comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown  
in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or a portion of this  
nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment  
encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule  
that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an  
20 integer between 1 and 61, is one that is sufficiently complementary to the nucleotide  
sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, that it can  
hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID  
NO:2*n*-1, wherein *n* is an integer between 1 and 61, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen  
25 base pairing between nucleotides units of a nucleic acid molecule, and the term “binding”  
means the physical or chemical interaction between two polypeptides or compounds or  
associated polypeptides or compounds or combinations thereof. Binding includes ionic,  
non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction  
can be either direct or indirect. Indirect interactions may be through or due to the effects of  
30 another polypeptide or compound. Direct binding refers to interactions that do not take  
place through, or due to, the effect of another polypeptide or compound, but instead are  
without other substantial chemical intermediates.

A “fragment” provided herein is defined as a sequence of at least 6 (contiguous)  
nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific



hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence.

Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

5           A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an  
10 in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

          A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a  
15 nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. A  
"homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is  
20 derived from different species.

          Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95%  
25 identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT  
30 PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

          A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those

sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

#### **NOVX Nucleic Acid and Polypeptide Variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically

result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1 and 61, are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human

NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1 and 61. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at

excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions

are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

## 10 Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61; still more

preferably at least about 80% homologous to SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1 and 61; even more preferably at least about 90% homologous to SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1 and 61; and most preferably at least about 95% homologous to SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1 and 61.

5           An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1 and 61, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1 and 61, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded  
10   protein.

          Mutations can be introduced any one of SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1 and 61, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid  
15   substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine,  
20   tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another  
25   embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2 $n$ -1, wherein  $n$  is an integer between 1 and 61, the encoded protein can be expressed by any recombinant technology known in the art and the  
30   activity of the protein can be determined.

          The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY,

FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single  
 5 letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and a NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an  
 10 intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (e.g., regulation of insulin release).

### **Antisense Nucleic Acids**

Another aspect of the invention pertains to isolated antisense nucleic acid molecules  
 15 that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or  
 20 complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61, or antisense  
 25 nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which  
 30 are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences



which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional  
5 nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and  
10 then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve  
15 sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units,  
20 the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

### **Ribozymes and PNA Moieties**

25 Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

30 In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in

Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, *e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. See, *e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. See, *e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial

restriction enzymes when used in combination with other enzymes, *e.g.*, S<sub>1</sub> nucleases (*See*, Hyrup, *et al.*, 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

**NOVX Polypeptides**

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61. The invention also includes a  
5 mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant  
10 in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a  
15 conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue  
20 sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion  
25 thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or  
30 recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about

10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between

1 and 61, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 61.

### Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent

identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

### Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a  
5 NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-  
linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide  
having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2 $n$ ,  
wherein  $n$  is an integer between 1 and 61, whereas a "non-NOVX polypeptide" refers to a  
polypeptide having an amino acid sequence corresponding to a protein that is not  
10 substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the  
NOVX protein and that is derived from the same or a different organism. Within a NOVX  
fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX  
protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-  
active portion of a NOVX protein. In another embodiment, a NOVX fusion protein  
15 comprises at least two biologically-active portions of a NOVX protein. In yet another  
embodiment, a NOVX fusion protein comprises at least three biologically-active portions  
of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to  
indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame  
with one another. The non-NOVX polypeptide can be fused to the N-terminus or  
20 C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the  
NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase)  
sequences. Such fusion proteins can facilitate the purification of recombinant NOVX  
polypeptides.

25 In another embodiment, the fusion protein is a NOVX protein containing a  
heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host  
cells), expression and/or secretion of NOVX can be increased through use of a  
heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion  
30 protein in which the NOVX sequences are fused to sequences derived from a member of  
the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the  
invention can be incorporated into pharmaceutical compositions and administered to a  
subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the



surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

#### NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of

limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

5 Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a  
10 variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods  
15 which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences.  
20 Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

### Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be  
25 used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-  
30 stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S<sub>1</sub> nuclease, and ligating the resulting fragment library into an expression vector. By this method,

expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

#### Anti-NOVX Antibodies

Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,  $F_{ab}$ ,  $F_{ab'}$  and  $F_{(ab')_2}$  fragments, and an  $F_{ab}$  expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes; subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the

invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2 $n$ , wherein  $n$  is an integer between 1 and 61, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant ( $K_D$ ) is  $\leq 1 \mu\text{M}$ , preferably  $\leq 100 \text{ nM}$ , more preferably  $\leq 10 \text{ nM}$ , and most preferably  $\leq 100 \text{ pM}$  to about  $1 \text{ pM}$ , as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

### **Polyclonal Antibodies**

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity

chromatography. Purification of immunoglobulins is discussed, for example, by D.

Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

### **Monoclonal Antibodies**

5           The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus  
10       contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

          Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an  
15       immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

          The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells  
20       of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly  
25       myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium  
30       for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

          Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a

medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also  
5 have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the  
10 binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard,  
15 Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable  
20 culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification  
25 procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*,  
30 by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce

immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

### Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).



### Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial

chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the

5 Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies.

10 Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S.

15 Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem  
20 cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in  
25 culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that  
30 binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

### **F<sub>ab</sub> Fragments and Single Chain Antibodies**

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab')<sub>2</sub></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an F<sub>(ab')<sub>2</sub></sub> fragment; (iii) an F<sub>ab</sub> fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F<sub>v</sub> fragments.

### **Bispecific Antibodies**

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH<sub>2</sub>, and CH<sub>3</sub> regions. It is preferred to have the first heavy-chain constant region (CH<sub>1</sub>) containing the site necessary for light-chain binding present in at

least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

5           According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced  
10       with larger side chains (*e.g.* tyrosine or tryptophan). Compensatory “cavities” of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

15           Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* F(ab')<sub>2</sub> bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to  
20       generate F(ab')<sub>2</sub> fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other  
25       Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

          Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')<sub>2</sub> molecule. Each  
30       Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$  domains of one fragment are forced to pair with the complementary  $V_L$  and  $V_H$  domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc $\gamma$ R), such as Fc $\gamma$ RI (CD64), Fc $\gamma$ RII (CD32) and Fc $\gamma$ RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

### **Heteroconjugate Antibodies**

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted

cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

### **Effector Function Engineering**

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

### **Immunoconjugates**

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolacca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, croton, saponaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the

tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., *Science*, **238**: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

### **Immunoliposomes**

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., *Proc. Natl. Acad. Sci. USA*, **82**: 3688 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA*, **77**: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., *J. Biol. Chem.*, **257**: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, *J. National Cancer Inst.*, **81**(19): 1484 (1989).

### **Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention**

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (*e.g.*, a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable



prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of  
5 bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### **Antibody Therapeutics**

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed  
10 to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance,  
15 administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

20 Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

25 A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding  
30 affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume of the subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1

mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

### Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other  
5 molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995;  
10 Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the  
15 antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant  
20 DNA technology. See, *e.g.*, Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a  
25 cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example,  
30 hydroxymethylcellulose or gelatin-microcapsules and poly-(methylemethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT<sup>TM</sup> (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

#### ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F<sub>ab</sub> or F<sub>(ab)2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an

analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

## 10 NO VX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NO VX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

30 The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the

nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the

fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.*, Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus

40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

5           In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477),  
10           pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).  
15           20

          The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is  
25           antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in  
30           the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub,

*et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).



A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

### Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, can be introduced as a transgene into

the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur

between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g., Thomas, et al., 1987. Cell 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. See, e.g., Li, et al., 1992. Cell 69: 915.*

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. Curr. Opin. Biotechnol. 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.*

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of Saccharomyces cerevisiae. See, O'Gorman, et al., 1991. Science 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.*

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, et al., 1997. Nature 385: 810-813. In brief, a cell (e.g., a somatic cell) from the transgenic animal can be isolated and induced to exit

the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

### Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates,

citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

5           Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be  
10       sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable  
15       mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be  
20       preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

          Sterile injectable solutions can be prepared by incorporating the active compound  
25       (e.g., a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the  
30       preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to

viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

## **Screening and Detection Methods**

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes

(regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

### Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or



biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 5 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. 10 *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; 15 Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast 20 cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>3</sup>H, either 25 directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound 30 form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially

bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , *etc.*), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and

determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>,

Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target

molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows

transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

5           The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

### Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

### 15           Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

30           Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes.

By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924.* Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see, Verma, et al., HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such

data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

### Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual



humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

### **Predictive Medicine**

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

### Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 61, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include

tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

### **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject

having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a

non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.,* U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.,* Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see, Abravaya, et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.,* genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see, Guatelli, et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see, Kwoh, et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q $\beta$  Replicase (*see, Lizardi, et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction

endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. See, e.g., Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded

duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S<sub>1</sub> nuclease to enzymatically digesting the mismatched regions. In other  
5       embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods*  
10       *Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations  
15       in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.,* a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test  
20       cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation  
25       polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of  
30       single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes

heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g., Keen, et al., 1991. Trends Genet. 7: 5.*

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing

5 gradient gel electrophoresis (DGGE). *See, e.g., Myers, et al., 1985. Nature 313: 495.*

When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of

high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is

used in place of a denaturing gradient to identify differences in the mobility of control and  
10 sample DNA. *See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.*

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions  
15 that permit hybridization only if a perfect match is found. *See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230.* Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective  
20 PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448*) or at the extreme 3'-terminus of one primer

where, under appropriate conditions, mismatch can prevent, or reduce polymerase  
25 extension (*see, e.g., Prossner, 1993. Tibtech. 11: 238*). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g., Gasparini, et al., 1992. Mol. Cell Probes 6: 1.* It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for

30 amplification. *See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189.* In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.



The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

### Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act

on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited  
5 enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic  
10 polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the  
15 extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive  
20 standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

25 Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness  
30 phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

### Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials.

5 For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels,  
10 or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

15 By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the  
20 levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a  
25 marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein,  
30 peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the

preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

### Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

### Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention

or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with  
5 Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or  
10 RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis,  
15 immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

### Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to  
20 the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a  
25 disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

### 30 Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention

involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX

5 peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX  
10 antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent  
15 identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is  
20 abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preeclampsia).

## 25 **Determination of the Biological Effect of the Therapeutic**

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with  
30 representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly,

for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

### Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

### Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

#### Example 1.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis		
	SEQ ID NO: 1	960 bp
NOV1a, CG103191-02	CTCGCCCCGGTGCCTAGGTGCCCGGCCACACCGCCAGCTGCTCGGCGCCCCGGGTCCG CCATGCGCTCCGCCGCTGTCCTGGCTCTTCTGCTCTGCGCCGGGCAAGTCACTGCGCT CCCTGTGAACAGCCCTATGAATAAAGGGGATACCGAGGTGATGAAATGCATCGTTGAG	

DNA Sequence	GTCATCTCCGACACACTTTCCAAGCCCAGCCCCATGCCTGCTCAGCCAGGAATGTTTTG AGACACTCCGAGGAGATGAACGGATCCTTTCCATTCTGAGACATCAGAATTTACTGAA GGAGCTCCAAGACCTCGCTCTCCAAGGCGCCAAGGAGAGGGCACATCAGCAGAAGAAA CACAGCGGTTTTTTGAAGATGAACTCTCAGAGGTTCTTGAGAACCAGAGCAGCCAGGCCG AGCTGAAAGAGGCGGTGGAAGAGCCATCATCCAAGGATGTTATGGAGAAAAGAGAGGA TTCCAAGGAGGCAGAGAAAAGTGGTGAAGCCACAGACGGAGCCAGGCCCCAGGCCCTC CCGGAGCCCATGCAGGAGTCCAAGGCTGAGGGGAACAATCAGGCCCTTGGGGAGGAAG AGGAGGAGGAGGAGGAGGCCACCAACACCCACCCTCCAGCCAGCCTCCCAGCCAGAA ATACCCAGGCCCACAGGCCGAGGGGGACAGTGAGGGCCTCTCTCAGGGTCTGGTGGAC AGAGAGAAGGGCCTGAGTGCAGAGCCCGGTGGCAGGCAAAGAGAGAAAGAGGAGGAGG AGGAGGAGGAGGCTGAGGCTGGAGAGGAGGCTGTCCCCGAGGAAGAAGGCCCTACTGT AGTGCTGAACCCCGAGGAGAAGAAAGAGGAGGAGGGCAGCGCAAACCGCAGACCAGAG GACCAGGAGCTGGAGAGCCTGTCTGGCCATTGAAGCAGAGCTGGAGAAAGTGGCCACC AGCTGCAGGCACTACGGCGGGGCTGAGACACC		
	ORF Start: ATG at 61	ORF Stop: TGA at 952	
	SEQ ID NO: 2	297 aa	MW at 32591.3 Da
NOV1a, CG103191-02 Protein Sequence	MRSAAVLALLLCAGQVTALPVNSPMNKGDTVMKCIVEVISDTLSKPSMPVVSQECFE TLRGDERILSILRHQNLKELQDLALQAKERAHQKKHSGFEDELSEVLENQSSQAE LKEAVEEPSSKDVMEKREDSKEAEKSGEATDGARPPALPEPMQESKAEGNNQAPGEEE EEEEEEATNTHPPASLPSQKYPGPQAEGDSEGLSQGLVDREKGLSAEPGWQAKREEEEE EEEEAEAGEEAVPEEEGPTVVLNPEEKKEEEGSANRRPEDQELESLSAIEAELEKVAHQ LQALRRG		
	SEQ ID NO: 3	837 bp	
NOV1b, CG103191-03 DNA Sequence	CCACACCGTCAGCTGCTCGGCGCCCGGGTCCGCCATGCGCTCCGCCGCTGTCTTGGCT CTTCTGCTCTGCGCCGGGCAAGTCACTGCGCTCCCTGTGAACAGCCCTATGAATAAAG GGGATACCGAGGTGATGAAATGCATCGTTGAGGTCATCTCCGACACACTTTCCAAGCC CAGCCCCATGCCTGTCTAGCCAGGAATGTTTTGAGACACTCCGAGGAGATGAACGGATC CTTTCCATTCTGAGACATCAGAATTTACTGAAGGAGCTCCAAGACCTCGCTCTCCAAG GCGCCAAGGAGAGGGGCACATCAGCAGAAGAAACACAGCGGTTTTGAAGATGAACTCTC AGAGGTTCTTGAGAACCAGAGCAGCCAGGCCGAGCTGAAAGAGGCGGTGGAAGAGCCA TCATCCAAGGATGTTATGGAGAAAAGAGAGGATTCCAAGGAGGCAGAGAAAAGTGGTG AAGCCACAGACGGAGCCAGGCCCCAGGCCCTCCCGGAGCCCATGCAGGACAACCGGGA CAGTTCCATGAAGCTCTCCTTCCGGGCCCGGGCCTACGGCTTCAGGGGCCCTGGGCCG CAGCTGCGACGAGGCTGGAGGCCATCCTCCTGGGAGGACAGCCTTGAGGCGGGCCTGC CCCTCCAGGTCCGAGGCTACCCCGAGGAGAAGAAAGAGGAGGAGGGCAGCGCAAACCG CAGACCAGAGGACCAGGAGCTGGAGAGCCTGTCTGGCCATTGAGGCAGAGCTGGAGAAA GTGGCCCAACAGCTGCGGGCACTACGGCGGGGCTGAGACACCGGCTGGCAGGGCTGGC CCAGGGCACCCCTGTGGGCCTGGCT		
	ORF Start: ATG at 35	ORF Stop: TGA at 788	
	SEQ ID NO: 4	251 aa	MW at 28029.1 Da
NOV1b, CG103191-03 Protein Sequence	MRSAAVLALLLCAGQVTALPVNSPMNKGDTVMKCIVEVISDTLSKPSMPVVSQECFE TLRGDERILSILRHQNLKELQDLALQAKERAHQKKHSGFEDELSEVLENQSSQAE LKEAVEEPSSKDVMEKREDSKEAEKSGEATDGARPPALPEPMQDNRDSSMKLSFRARA YGFRGPGPQLRRGWRPSSWEDSLEAGLPLOVRGYPEEKKEEEGSANRRPEDQELESLS AIEAELEKVAHQLRALRRG		
	SEQ ID NO: 5	1002 bp	
NOV1c, CG103191-04 DNA Sequence	CCACACCGCCAGCTGCTCGGCGCCCGGGTCCGCCATGCGCTCCGCCGCTGTCTTGGCT CTTCTGCTCTGCGCCGGGCAAGTCACTGCGCTCCCTGTGAACAGCCCTATGAATAAAG GGGATACCGAGGTGATGAAATGCATCGTTGAGGTCATCTCCGACACACTTTCCAAGCC CAGCCCCATGCCTGTCTAGCCAGGAATGTTTTGAGACACTCCGAGGAGATGAACGGATC CTTTCCATTCTGAGACATCAGAATTTACTGAAGGAGCTCCAAGACCTCGCTCTCCAAG GCGCCAAGGAGAGGGGCACATCAGCAGAAGAAACACAGCGGTTTTGAAGATGAACTCTC AGAGGTTCTTGAGAACCAGAGCAGCCAGGCCGAGCTGAAAGGTCGGTCGGAGGCTCTG GCTGTGGATGGAGCTGGGAAGCCTGGGGCTGAGGAGGCTCAGGACCCCGAAGGGAAGG GAGAACAGGAGCACTCCAGCAGAAAGAGGAGGAGGAGGAGATGGCAGTGGTCCCGCA		



	AGGCCTCTTCCGGGGTGGGAAGAGCGGAGAGCTGGAGCAGGAGGAGGAGCGGCTCTCC AAGGAGTGGGAGGACTCCAAACGCTGGAGCAAGATGGACCAGCTGGCCAAGGAGCTGA CGGCTGAGAAGCGGCTGGAGGGGCAGGAGGAGGAGGAGGACAACCGGGACAGTTCAT GAAGCTCTCCTTCCGGGGCCCGGGCTACGGCTTCAGGGGGCCTGGGCCGAGCTGCGA CGAGGCTGGAGGCCATCCTCCCGGGAGGACAGCCTTGAGGCGGGCCTGCCCCCTCCAGG TCCGAGGCTACCCCGAGGAGAAGAAAGAGGAGGAGGGCAGCGCAAACCGCAGACCAGA GGACCAGGAGCTGGAGAGCCTGTTCGGCCATTGAGGCAGAGCTGGAGAAAGTGGCCAC CAGCTGCAGGCACTACGGCGGGGCTGAGACACCGGCTGGCAGGGCTGGCCCCAGGGCA CCCTGTGGGCTGGCT		
	ORF Start: ATG at 35	ORF Stop: TGA at 953	
	SEQ ID NO: 6	306 aa	MW at 34268.8 Da
NOV1c, CG103191-04 Protein Sequence	MRSAAVLALLLCAGQVTALPVNSPMNKGDTVMKCIVEVISDTLSKPSPMPVVSQECFE TLRGDERILSILRHQNLKELQDLALQGAKERAHQKKHSGFEDELSEVLENQSSQAE LKGRSEALAVDGAGKPGAEEAQDPEGKGEQEHSSQKEEEEEEMAVVPQGLFRGKSGEL EQEEERLSKEWEDSKRWSKMDQLAKELTAEKRLLEGQEEEDNRDSSMKLSFRARAYGF RGPQPQLRRGWRPSSREDSLEAGLPLQVRGYPEEKKEEEGSANRRRPEDQELESLSAIE AELEKVAHQQLALRRG		
	SEQ ID NO: 7	337 bp	
NOV1d, 251425133 DNA Sequence	CACCAGATCTCTCCCTGTGAACAGCCCTATGAATAAAGGGGATACCGAGGTGATGAAA TGCATCGTTGAGGTCATCTCCGACACACTTTCCAAGCCCAGCCCCATGCCTGTACGCC AGGAATGTTTTGAGACACTCCGAGGAGATGAACGGATCCTTTCCATTCTGAGACATCA GAATTTACTGAAGGAGCTCCAAGACCTCGCTCTCCAAGCGCCAAGGAGAGGGCACAT CAGCAGAAGAAACACAGCGGTTTTGAAGATGAACCTCAGAGGTTCTTGAGAACCCAGA GCAGCCAGGCCGAGCTGAAAGGTCGGTCGGAGGCTCTGCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 8	112 aa	MW at 12528.0 Da
NOV1d, 251425133 Protein Sequence	TRSLPVNSPMNKGDTVMKCIVEVISDTLSKPSPMPVVSQECFETLRGDERILSILRHQ NLKELQDLALQGAKERAHQKKHSGFEDELSEVLENQSSQAELEKGRSEALLEG		
	SEQ ID NO: 9	595 bp	
NOV1e, 251425611 DNA Sequence	CACCAGATCTGCCGAGCTGAAAGGTCGGTCGGAGGCTCTGGCTGTGGATGGAGCTGGG AAGCCTGGGGCTGAGGAGGCTCAGGACCCGAAGGGAAGGGAGAACAGGAGCACTCCC AGCAGAAAGAGGAGGAGGAGGAGATGGCAGTGGTCCCGAAGGCCTCTTCCGGGGTGG GAAGAGCGGAGAGCTGGAGCAGGAGGAGGAGCGGCTCTCCAAGGAGTGGGAGGACTCC AAACGCTGGAGCAAGATGGACCAGCTGGCCAAGGAGCTGACGGCTGAGAAGCGGCTGG AGGGGCAGGAGGAGGAGGAGGACAACCGGGACAGTTCCATGAAGCTCTCCTTCCGGGC CCGGGCCCTACGGCTTCAGGGGGCCTGGGCCGAGCTGCGACGAGGCTGGAGGCCATCC TCCCGGGAGGACAGCCTTGAGGCGGGCCTGCCCCCTCCAGGTCCGAGGCTACCCGAGG AGAAGAAAGAGGAGGAGGGCAGCGCAAACCGCAGACCAGAGGACCAGGAGCTGGAGAG CCTGTCTGGCCATTGAGCGGAGCTGGAGAAAGTGGCCCACCAGCTGCAGGCACTACGG CGGGCCCTCAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 10	198 aa	MW at 22331.2 Da
NOV1e, 251425611 Protein Sequence	TRSAELKGRSEALAVDGAGKPGAEEAQDPEGKGEQEHSSQKEEEEEEMAVVPQGLFRGG KSGELEQEEERLSKEWEDSKRWSKMDQLAKELTAEKRLLEGQEEEDNRDSSMKLSFRA RAYGFRGPGPQLRRGWRPSSREDSLEAGLPLQVRGYPEEKKEEEGSANRRRPEDQELES LSAIEAELEKVAHQQLALRRGLEG		
	SEQ ID NO: 11	718 bp	
NOV1f, 278460276 DNA Sequence	CACCAGATCTCTCCCTGTGAACAGCCCTATGAATAAAGGGGATACCGAGGTGATGAAA TGCATCGTTGAGGTCATCTCCGACACACTTTCCAAGCCCAGCCCCATGCCTGTACGCC AGGAATGTTTTGAGACACTCCGAGGAGATGAACGGATCCTTTCCATTCTGAGACATCA		

	GAATTTACTGAAGGAGCTCCAAGACCTCGCTCTCCAAGGCGCCAAGGAGAGGGGCACAT CAGCAGAAGAAACACAGCGTTTTTGAAGATGAACTCTCAGAGGTTCTTGAGAACCAGA GCAGCCAGGCCGAGCTGAAAGAGGCGGTGGAAGAGCCATCATCCAAGGATGTTATGGA GAAAAGAGAGGATTCCAAGGAGGCAGAGAAAAGTGGTGAAGCCACAGACGGAGCCAGG CCCCAGGCCCTCCCGGAGCCCATGCAGGACAACCGGGACAGTTCCATGAAGCTCTCCT TCCGGGCCCGGGCCTACGGCTTCAGGGGCCCTGGGCGCAGCTGCGACGAGGCTGGAG GCCATCCTCCTGGGAGGACAGCCTTGAGGCGGGCCTGCCCCCAGGTCCGAGGCTAC CCCGAGGAGAAGAAAAGAGGAGGAGGGCAGCGCAAACCGCAGACCAGAGGACCAGGAGC TGGAGAGCCTGTTCGGCCATTGAGGCAGAGCTGGAGAAAAGTGGCCCACCAGCTGCGGGC ACTACGGCGGGGCCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 12	239 aa	MW at 26902.7 Da
NOV1f, 278460276 Protein Sequence	TRSLPVNSPMNKGDTEVMKCIVEVISDTLSKPSMPVVSQECFETLRGDERILSILRHQ NLLKELQDLALQGAKEAHQKKHSGFEDELSEVLENQSSQAELEAVEEPSSKDVME KREDSKEAEKSGEATDGAHPQALPEPMQDNRRSSMKLSFRARAYGFRGPGPQLRRGWR PSSWEDSLEAGLPLQVRGYPEEKKEEESANRRPEDQELESLSAIEAELEKVAHQQLRA LRRGLEG		
	SEQ ID NO: 13	856 bp	
NOV1g, 278456175 DNA Sequence	CACCAGATCTCTCCCTGTGAACAGCCCTATGAATAAAGGGGATACCGAGGTGATGAAA TGCATCGTTGAGGTCATCTCCGACACACTTTCCAAGCCCAGCCCCATGCCTGTCAGCC AGGAATGTTTTGAGACACTCCGAGGAGATGAACGGATCCTTTCCATTCTGAGACATCA GAATTTACTGAAGGAGCTCCAAGACCTCGCTCTCCAAGGCGCCAAGGAGAGGGGCACAT CAGCAGAAGAAACACAGCGTTTTTGAAGATGAACTCTCAGAGGTTCTTGAGAACCAGA GCAGCCAGGCCGAGCTGAAAGAGGCGGTGGAAGAGCCATCATCCAAGGATGTTATGGA GAAAAGAGAGGATTCCAAGGAGGCAGAGAAAAGTGGTGAAGCCACAGACGGAGCCAGG CCCCAGGCCCTCCCGGAGCCCATGCAGGAGTCCAAGGCTGAGGGGAACAATCAGGCC CTGGGGAGGAAGAGGAGGAGGAGGAGGAGGCCACCAACACCCACCCTCCAGCCAGCCT CCCCAGCCAGAAATACCCAGGCCACAGGCCGAGGGGGACAGTGAGGGCCTCTCTCAG GGTCTGGTGGACAGAGAGAAGGGCCTGAGTGCAGAGCCCGGTGGCAGGCAAAGAGAG AAGAGGAGGAGGAGGAGGAGGAGGCTGAGGCTGGAGAGGAGGCTGTCCCCGAGGAAGA AGGCCCCACTGTAGTGCTGAACCCCGAGGAGAAGAAAGAGGAGGAGGGCAGCGCAAAC CGCAGACCAGAGGACCAGGAGCTGGAGAGCCTGTTCGGCCATTGAAGCAGAGCTGGAGA AAGTGGCCCACCAGCTGCAGGCACTACGGCGGGGCCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 14	285 aa	MW at 31464.9 Da
NOV1g, 278456175 Protein Sequence	TRSLPVNSPMNKGDTEVMKCIVEVISDTLSKPSMPVVSQECFETLRGDERILSILRHQ NLLKELQDLALQGAKEAHQKKHSGFEDELSEVLENQSSQAELEAVEEPSSKDVME KREDSKEAEKSGEATDGAHPQALPEPMQESKAEGNNQAPGEEEEEEATNTHPPASL PSQKYPGPQAEGLDSEGLSQGLVDREKGLSAEPGWQAKREEEEEEAEAGEEAVPEEE GPTVVLNPEEKKEEESANRRPEDQELESLSAIEAELEKVAHQQLALRRGLEG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

Table 1B. Comparison of NOV1a against NOV1b through NOV1g.		
Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV1b	1..297 1..251	201/297 (67%) 212/297 (70%)
NOV1c	1..297 1..306	172/313 (54%) 188/313 (59%)

NOVId	18..118 3..103	100/101 (99%) 101/101 (99%)
NOVle	192..297 94..195	46/109 (42%) 55/109 (50%)
NOVIf	18..297 3..236	183/280 (65%) 195/280 (69%)
NOVlg	18..297 3..282	236/280 (84%) 237/280 (84%)

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

Table 1C. Protein Sequence Properties NOV1a	
PSort analysis:	0.7618 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1D.

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY53797	Amino acid sequence of the mature human chromogranin A (CgA) protein - Homo sapiens, 439 aa. [WO9958980-A1, 18-NOV-1999]	19..255 1..238	237/238 (99%) 237/238 (99%)	e-132
AAU86000	Modified vasostatin II antibiotic peptide - Unidentified, 113 aa. [WO200210195-A2, 07-FEB-2002]	19..131 1..113	113/113 (100%) 113/113 (100%)	2e-58
AAY53798	Amino acids 145-234 of the mature human chromogranin A (CgA) protein - Homo sapiens, 90 aa. [WO9958980-A1, 18-NOV-1999]	163..251 1..90	89/90 (98%) 89/90 (98%)	4e-45
AAB37069	Recombinant vasostatin I peptide - Unidentified, 81 aa. [FR2792638-A1, 27-OCT-2000]	17..96 2..81	80/80 (100%) 80/80 (100%)	2e-39
AAB37066	Human vasostatin I peptide - Homo sapiens, 76 aa. [FR2792638-A1, 27-	19..94 1..76	76/76 (100%) 76/76 (100%)	4e-37

	OCT-2000]			
--	-----------	--	--	--

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

<b>Table 1E. Public BLASTP Results for NOV1a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV1a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
A28468	chromogranin A precursor [validated] - human, 457 aa.	1..255 1..256	255/256 (99%) 255/256 (99%)	e-142
P10645	Chromogranin A precursor (CGA) (Pituitary secretory protein I) (SP-I) [Contains: Vasostatin I; Vasostatin II; EA-92; ES-43; Pancreastatin SS-18; WA-8; WE-14; LF-19; AL-11; GV-19; GR-44; ER-37] - Homo sapiens (Human), 457 aa.	1..255 1..256	255/256 (99%) 255/256 (99%)	e-142
Q96GL7	Similar to chromogranin A (Parathyroid secretory protein I) - Homo sapiens (Human), 407 aa (fragment).	54..255 4..206	202/203 (99%) 202/203 (99%)	e-111
P05059	Chromogranin A precursor (CGA) (Pituitary secretory protein I) (SP-I) [Contains: Vasostatin-I; Chromostatin; Chromacin; Pancreastatin; WE- 14; Catestatin] - Bos taurus (Bovine), 449 aa.	1..271 1..273	202/276 (73%) 215/276 (77%)	e-100
A41520	chromogranin A precursor [validated] - bovine, 449 aa.	1..271 1..273	199/276 (72%) 213/276 (77%)	3e-99

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1F.

<b>Table 1F. Domain Analysis of NOV1a</b>			
<b>Pfam Domain</b>	<b>NOV1a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
Granin	1..297	138/689 (20%) 291/689 (42%)	1.7e-29

5

### Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

<b>Table 2A. NOV2 Sequence Analysis</b>		
	SEQ ID NO: 15	2521 bp

NOV2a, CG105757-01 DNA Sequence	ACAGTTGTAAGGGATCTTGTGGCTGTCAGGATGGCAGAGGAGCAGGAGTTCACCCAGC TCTGCAAGTTGCCGTGCACAGCCCTCACACCCACACTGCGTGAACAACACCTACCGCAG CGCACAGCACTCCCAGGCTCTGCTCCGAGGGCTGCTGGCTCTCCGGGACAGCGGAATC CTCTTCGATGTTGTGCTGGTGGTGGAGGGCAGACACATCGAGGCCATCGCATCCTGC TGGCTGCGTCTCTGCGATTACTTCAGGAGAGGAATGTTTGCTGGGGGATTGAAGGAGAT GGAACAGGAAGAGGTCTCTGATCCACGGTGTGCTTACAATGCTATGTGCCAAATCCTA CATTTTCATATACACCTCCGAGCTGGAGCTCAGCCTGAGCAATGTACAAGAGACACTGG TGGCTGCCGTGCCAGCTGCAGATCCCAGAAATTATCCATTCTGCTGTGATTTCTCTCAT GTCCTGGGTGGACGAAGAGAACATTCTCGATGTCTACCGGTGGCAGAGCTGTTTGAC TTGAGCCGCCCTGACTGAGCAACTGGACACCTATATCCTCAAAAACCTTTGTGGCCTTCT CTCGGACTGACAAGTACCGCCAGCTTCCATTGGAGAAGGTCTACTCCCTCCTCAGCAG CAATCGCCTGGAGGTCTCTGCGAGACCGAGGTATATGAGGGGGCCCTTCTCTACCAT TATAGCCTGGAGCAGGTGCAGGCTGACCAGATCTCGCTGCACGAGCCCCCAAAGCTCC TTGAGACAGTGCGGTTTCCGCTGATGGAAGCTGAGGTCTGACGCGGTGCATGACAA GCTGGACCCAGCCCTTTGAGGGACACAGTGGCCAGCGCCCTCATGTACCACCGGAAC GAGAGCCTACAGCCAGCCTGCAGAGCCCGCAAACGGAGCTGCGGTGCGACTTCCAGT GCGTTGTGGGCTTCGGGGGCATTCACTCCACGCGCTCCACTGTCTCAGCGACCAAGC CAAGTATCTAAACCCCTTACTGGGAGAGTGGAAAGCACTTCACTGCCCTCCCTGGCCCC CGCATGTCCAACCAGGGCATCGCGGTGCTCAACAACCTTCGTATACCTTGATTGGAGGGG ACAACAATGTCCAAGGATTTTCGAGCAGAGTCCCGATGCTGGAGGTATGACCCACGGCA CAACCGCTGGTTCCAGATCCAGTCCCTGCAGCAGGAGCACGCGGACCTGTCCGTGTGT GTTGTAGGCAGGTACATCTACGCTGTGGCGGGCCGTGACTACCACAATGACCTGAATG CTGTGGAGCGCTACGACCCTGCCACCAACTCCTGGGCATACGTGGCCCCACTCAAGAG GGAGGTAGTGTATGCCACGCAGGCGCGACGCTGGAGGGGAAGATGTATATCACCTGC GGCCGCAGAGGGGAGGATTACCTGAAAGAGACACACTGCTACGATCCAGGCAGCAACA CTTGGCACACACTGGCTGATGGGCCTGTGCGGCGCGCCTGGCACGGCATGGCAACCCT CCTCAACAAGCTGTATGTGATCGGGGGCAGCAACAACGATGCCGGATACAGGAGGGAC GTGCACCAGCTCCAGGTGCCACGTGCTGCGCTGGCTGGAGGCAGCAAGGGGACGAG TGTGGGATTTGCGGTGTGCGAAGGCAACTCCACGTGCTCAGAGAACGAGGTGTGCGTGA GGCCTGGCGAGTGCCGCTGCCGCCACGGCTACTTCGGTGCCAACTGCGACACCAAGTGT GGCCAGTGCAAGGGGGCCAGCAGCCGTGCACGGTGGCCGAGGGCCGCTGCTTGACGTGC GAGCCCGGTGGAACGGAACCAAGTGCGACCAAGCCTTGCGCCACCGGTTTCTATGGCG AGGGCTGCAGCCACCGCTGTCCGCCATGCCGCGACGGGCATGCCCTGTAACCATGTAC CGGCAAGTGTACGCGCTGCAACGCGGGCTGGATCGGCGACCGGTGCGAGACCAAGTGT AGCAATGGCACTTACGGCGAGGACTGCGCCTTCGTGTGCGCCGACTGCGGCAGCGGAC ACTGCGACTTCCAGTTCGGGGCGCTGCTGCTGTCAGCCCTGGCGTCCAGGGGCCCCAGTG AGTGCCCCGGGACCGGGAGGGGTTGGGGCTTGTACCTGCCACAGAGGGGGGTCCAG CCGACGAGGTGGCCTCTCCACCCTGAGCTGGGTTATCACCTCAGCCTTGGTCCCTTAC CCCAGCTAGGGAGTGACAGTAGGCTCTTTGGGGCAGTTTCTGCTGGATGTGCGGG AGCTCACGTTTACGCGCAGGATCTGGTGACCAGTCCAGCCTGTGTAGTGGGCTCTTAA GGTGACCCCGAGTTGGTACAGAAGGACCAGGGACCTCCACTTACAGCCAAGGGTCTGG TTCAGCAGCCCCCTCTTCCCACCTAGCCGAGTCAGCCCCAGCAGTGGGCGCTGCGGCGC GGCCACCACGGGTCTATCCCCAGGCCCCCCCACTAGTGTTGTGCAACATTCTGTTTC CAAAACATCCACTACCCAATATGTGCC
	ORF Start: ATG at 31 ORF Stop: TAA at 1903
	SEQ ID NO: 16 624 aa MW at 71369.7 Da
NOV2a, CG105757-01 Protein Sequence	MAEEQEFTQLCKLPAQPSHPHCVNNTYRSAQHSQALLRGLLALRDSGILFDVVLVVEG RHIEAHRILLAASCDYFRRGMFAGGLKEMEQUEEVLIHGVSYNAMCQILHFIYTSELEL SLSNVQETLVAACQLQIPEIIHFCCDFLMSWVDEENILDVYRLAELFDLSRLTEQLDT YILKNFVAFSRDQYRQLPLEKVYSLSSNRLEVSCETEVEGALLYHYSLEQVQADQ ISLHEPPKLETVRFPLMEAEVLQRLHDKLDPSPLRDTVASALMYHRNESLQPSLQSP QTELRSDFCVVGFGGIHSTPSTVLSDQAKYLNPLLGEWKHFTASLAPRMSNQGIAVL NNFVYLIGDNNVQGFRAESRCWRYDPRHNRWFQIQSLQOEHADLSVCVVGRIYIYAVA GRDYHNDLNAVERYDPATNSWAYVAPLKREVVYAHAGATLEGKMYITCGRRGEDYLKE THCYDPGSNTWHTLADGPVRRWHGMATLLNKLYVIGGSNNNDAGYRRDVHQLPGAHLV RWLEAARGRVWDCGVRRLHVLRRERGVREAWRVPLPPRLRLCQLRHQCQCKGPAAVH GGRGPLLDVRARLERNQVRPALRHRFLWRGLQPPLSAMPRRACL

Further analysis of the NOV2a protein yielded the following properties shown in Table 2B.

<b>Table 2B. Protein Sequence Properties NOV2a</b>	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2C.

<b>Table 2C. Geneseq Results for NOV2a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV2a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAM39985	Human polypeptide SEQ ID NO 3130 - Homo sapiens, 634 aa. [WO200153312-A1, 26-JUL-2001]	1..516 1..514	512/516 (99%) 514/516 (99%)	0.0
AAB92457	Human protein sequence SEQ ID NO:10499 - Homo sapiens, 525 aa. [EP1074617-A2, 07-FEB-2001]	1..503 1..501	501/503 (99%) 501/503 (99%)	0.0
AAB60095	Human transport protein TPPT-15 - Homo sapiens, 462 aa. [WO200078953-A2, 28-DEC-2000]	1..457 1..455	453/457 (99%) 454/457 (99%)	0.0
AAM41771	Human polypeptide SEQ ID NO 6702 - Homo sapiens, 524 aa. [WO200153312-A1, 26-JUL-2001]	1..458 66..521	447/458 (97%) 451/458 (97%)	0.0
ABG27028	Novel human diagnostic protein #27019 - Homo sapiens, 421 aa. [WO200175067-A2, 11-OCT-2001]	78..372 127..421	295/295 (100%) 295/295 (100%)	e-171

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2D.

<b>Table 2D. Public BLASTP Results for NOV2a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV2a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q96B68	Hypothetical 71.7 kDa protein - Homo	1..516	513/516 (99%)	0.0

	sapiens (Human), 634 aa.	1..514	514/516 (99%)	
Q96KC6	CDNA FLJ14360 fis, clone HEMBA1000488, weakly similar to RING CANAL protein - Homo sapiens (Human), 525 aa.	1..503 1..501	501/503 (99%) 501/503 (99%)	0.0
Q99JN2	Hypothetical 71.7 kDa protein - Mus musculus (Mouse), 634 aa.	1..516 1..514	487/516 (94%) 502/516 (96%)	0.0
Q96QI7	Hypothetical 68.2 kDa protein - Homo sapiens (Human), 604 aa.	27..504 18..493	170/486 (34%) 274/486 (55%)	4e-75
Q9P2N7	Hypothetical protein KIAA1309 - Homo sapiens (Human), 639 aa (fragment).	27..504 53..528	170/486 (34%) 274/486 (55%)	4e-75

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2E.

Table 2E. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
BTB	34..146	37/144 (26%) 87/144 (60%)	1.6e-22
Kelch	339..387	13/49 (27%) 37/49 (76%)	1.5e-06
Kelch	389..434	12/47 (26%) 35/47 (74%)	8e-07
Kelch	437..482	12/47 (26%) 31/47 (66%)	0.0079

### Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis		
	SEQ ID NO: 17	5369 bp
NOV3a, CG108175-01 DNA Sequence	ATCCTCTCCGGGCTGTTCCCTGGCCTGTCTGCTCCTCCGGGCTCTGTCCCAGCAGCGA CAATGAGCTCCACACTCCACTCGGTTTTCTTCACCCTGAAGGTCAGCATCCTGCTGGG GTCCCTGCTGGGGCTCTGCCTGGGCCTTGAGTTCATGGGCTCCCAACAGTGGGCC CGCTACCTCCGCTGGGATGCCAGCACACGCAGTGACCTGAGTTCCAGTTCAAGACCA ACGTCTCTACGGGGCTGCTCCTCTACCTGGATGATGGCGGCGTCTGCGACTTCCTATG CCTCTCCCTGGTGGATGGCCGCGTTTCAGCTCCGCTTCAGCATGGACTGTGCCGAGACT GCCGTGCTGTCCAACAAGCAGGTGAATGACAGCAGCTGGCACTTCCTCATGGTGAGCC GTGACCGCTGCGCACGGTGTGATGCTTGATGGCGAGGGCCAGTCTGGGGAGCTGCA GCCCCAGCGGCCCTACATGGATGTGGTCAGTGACTTGTTCCTTGGTGGAGTCCCTACT GACATACGACCTTCTGCCCTGACCCTTGATGGAGTTCAGGCCATGCCCGGCTTCAAGG GGTTAATTCTGGATCTCAAGTATGGAACTCGGAGCCTCGGCTTCTGGGGAGCCGGGG	

TGTCCAGATGGATGCCGAGGGACCCTGTGGTGGAGCGTCCCTGTGAAAATGGTGGGATC TGCTTTCTCCTGGACGGCCACCCACCTGTGACTGTTCTACCCTGGCTATGGTGGCA AGCTCTGCTCAGAAGGCCTCTCCACCTCATGATGAGTGAACAAGCTCGAGAGGAGAA TGTGGCCACTTTCCGAGGCTCAGAGTATCTGTGCTACGACCTGTCTCAGAACCCGATC CAGAGCAGCAGTGATGAAATCACCTCTCCTTTAAGACCTGGCAGCGTAACGGCCTCA TCCTGCACACGGGCAAGTCGGCTGACTATGTCAACCTGGCTCTGAAGGATGGTGCAGG CTCCTTGGTCATTAACTGGGGTCCGGGGCCTTTGAGGCCATTGTGGAGCCAGTGAAT GGAAAATTCAACGACAACGCCTGGCATGATGTCAAAGTGACACGCAACCTCCGGCAGG TGACAATCTCTGTGGATGGCATTCTTACCACGACGGGCTACACTCAAGAGGACTATAC CATGCTGGGCTCGGACGACTTCTTCTATGTAGGAGGAAGCCCAAGTACCGCTGACTTG CCTGGCTCCCCGTGTCAGCAACAACCTTCATGGGCTGCCTTAAAGAGGTTGTTTATAAGA ATAATGACATCCGTCTGGAGCTGTCTCGCCTGGCCCGGATTGCGGACACCAAGATGAA AATCTATGGCGAAGTTGTGTTTAAAGTGTGAGAATGTGGCCACACTGGACCCCATCAAC TTTGAGACCCAGAGGCTTACATCAGCTTGCCCCAAGTGGAACACTAAACGTATGGGCT CCATCTCCTTTGACTTCCGCACCACAGAGCCCAATGGCCTGATCCTCTTCACTCATGG AAAGCCCCAAGAGAGGAAGGATGCTCGGAGCCAGAAGAATACAAAAGTAGACTTCTTT GCCGTGGAATCCTTCGATGGCAACCTGTACTTGCTGCTTGACATGGGCTCTGGCACA TCAAAGTGAAAGCCACTCAGAAGAAAGCCCAATGATGGGGAATGGTACCATGTGGACAT TCAGCGAGATGGCAGATCAGGTACTATATCAGTGAACAGCAGGCGCACGCCATTACACC GCCAGTGGGGAGAGCGAGATCCTGGACCTGGAAGGAGACATGTACCTGGGAGGGCTGC CGGAGAACCGTGTCTGGCCTTATTCTCCCCACCGAGCTGTGGACTGCCATGCTCAACTA TGGCTACGTGGGCTGCATCCGCGACCTATTCAATTGATGGGCGCAGCAAGAACATTCTGA CAGCTGGCAGAGATGCAGAATGCTGCGGGTGTCAAGTCCCTCCTGTTCACGGATGAGTG CCAAGCAGTGTGACAGCTACCCCTGCAAGAATAATGCTGTGTGCAAGGACGGCTGGAA CCGCTTCATCTGCGACTGCACCGGCACCGGATACTGGGGAAGAACCTGCCAAAGGGAG GCATCCATCCTGAGCTATGATGGTAGCATGTACATGAAGATCATGCCCATGGTCA TGCATACTGAGGCAGAGGATGTGTCCTTCCGCTTCATGTCCAGCGAGCTTATGGGCT GCTGGTGGCTACGACCTCCAGGACTCTGCCGACACCCTGCGTCTGGAGCTGGATGGG GGGCGTGTCAAGCTCATGGTTAACTTAGACTGTATCAGGATAAACTGTAACCTCCAGCA AAGGACCAGAGACCTTGTATGCAGGGCAGAAGCTCAATGACAACGAGTGGCACACCCT TCGGGTGGTGCGGAGAGGAAAAAGCCTTAAGTTAACCGTGGATGATGATGTGGCTGAG GGTACAATGGTGGGAGACCATAACCGTTTGGAGTTCCACAACATTGAAACGGGAATCA TGACTGAGAAACGCTACATCTCCGTTGTCCCCTCCAGCTTTATTGGCCATCTGCAGAG CCTCATGTTTAAATGGCCTTCTCTACATTGACTTGTGCAAAAATGGTGACATTGATTAT TGTGAGCTGAAGGCTCGTTTTTGGACTGAGGAACATCATCGCTGACCTGTACCTTTA AGACCAAGAGCAGCTACCTGAGCCTTGCCACTCTTCAGGCTTACACCTCCATGCACCT CTTCTTCCAGTTCAAGACCACCTCACCAGATGGCTTCATTCTCTTCAATAGTGGTGAT GGCAATGACTTCATTGCAGTCGAGCTTGTCAAGGGGTATATACACTACGTTTTTGAAC TCGGAACCGTCCCAATGTGATCAAAGGCAACAGTGACCGCCCCCTGAATGACAACCA GTGGCACAATGTGTCATCACTCGGGACAATAGTAACACTCATAGCCTGAAAGTGGAC ACCAAAGTGGTCACTCAGGTTATCAATGGTGCCAAAAATCTGGATTTGAAAGGTGATC TCTATATGGCTGGTCTGGCCCAAGGCATGTACAGCAACCTCCCAAAGCTCGTGGCCTC TCGAGATGGCTTTTCAAGGCTGTCTAGCATCAGTGGACTTGAATGGACGCTGGCAGAC CTCATCAATGATGCTCTTTCATCGGAGCGGACAGATCGAGCTGGCTGTGAAGGTACAA CCTTACTAGGACCCAGTACCACCTGCCAGGAAGATTTCATGTGCCAACAGGGGGTCTG CATGCAACAATGGGAGGGCTTCACCTGTGATTGTTCTATGACCTCTTATTCTGGAAC CAGTGCAATGATCCTGGCGCTACGTACATCTTTGGGAAAAGTGGTGGGCTTATCCTCT ACACCTGGCCAGCCAATGACAGGCCAGCACGCGGTCTGACCGCCTTGCCGTGGGCTT CAGCACCCTGTGAAGGATGGCATCTTGGTCCGCATCGACAGTGCTCCAGGACTTGGT GACTTCCTCCAGCTTCACATAGAACAGGGGAAAATTGGAGTTGTCTTCAACATTGGCA CAGTTGACATCTCCATCAAAGAGGAGAGAACCCCTGTAAATGACGGCAAATACCATGT GGTACGCTTCACAGGAACGGCGGCAACGCCACCCTGCAGGTGGACAATGGCCAGTG AATGAACATTATCCTACAGGCAACACTGATAATGAACGCTTCCAAATGGTAAAACAGA AAATCCCCCTTCAAATATAATCGGCCTGTAGAGGAGTGGCTGCAGGAAAAAGGCCGGCA GTTAACCATCTTCAACACTCAGGCGCAAATAGCCATTGGTGGAAAGGACAAAGGACGC CTCTTCAAGGCCAACTCTCTGGGCTCTATTATGATGGTTTGAAGTACTGAACATGG CGGCTGAGAACAAACCCCAATATTAAATCAATGGAAGTGTTCGGCTGGTTGGAGAAGT CCCATCAATTTTGGGAACAACACAGACGACCTCCATGCCACCAGAAATGTCTACTACT GTCATGGAAACCACTACTACAATGGCGACTACCACAACCCGTAAGAATCGCTCTACAG CCAGCATTCAGCCAACATCAGATGATCTTGTTCATCTGCTGAATGTTCAAGTGTGA TGAAGACTTTGTTGAATGTGAGCCGAGTACAGGAGGTGAATTAGTTATCCCTCTTCTT
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



	GTAGAAGACCCTTTAGCTACCCCTCCTATTGCTACTCGTGACCTTCCATTACACTCC CCCCTACCTTTTCGCCCCCTCCTCACCATTATTGAGACCACCAAAGATTCCCTGTCCAT GACCTCTGAGGCGGGTTACCTTGCTTGTCGGACCAAGGCAGCGATGGTTGTGATGAT GATGGCTTGGTGATATCTGGGTATGGCTCAGGGGAAACCTTTGACTCTAACCTGCCCC CTACTGATGATGAAGATTTTTTACACCACCTTCTCCTTGGTAACAGATAAGAGTCTTTC CACTTCAATCTTTCGAAGGTGGCTACAAAGCACATGCGCCCAAGTGGGAATCCAAGGAC TTTAGACCTAACAAAGTCTCCGAAACTAGTAGGACTACTACCACATCTTTATCCCCTG AGCTGATCCGCTTCACAGCTTCTCCTCCTCGTCTGGGATGGTGCCCAAATTGCCAGCTGG CAAAATGAATAACCGTGATCTCAAACCCACGCTGATATAGTCTTGCTTCCGTTGCC ACTGCCTATGAGCTAGACAGCACCAAACCTGAAGAGCCCACTAATTACTTCCCCCATGT TCCGTAATGTGCCCACAGCAAACCCACGGAGCCGGGAATCAGACGGGTTCGGGGGG CTCAGAGGTGATCCGGGAGTCGAGCAGCACAAACAGGGATGGTCGTCGGCATTGTGGCT GCTGCCGCCCTCTGCATCTTGATCCTCCTGTACGCCATGTACAAGTACAGGAACAGGG ACGAGGGGTCTTATCAAGTGGACGAGACGCGGAACCTACATCAGCAACTCCGCCAGAG CAACGGCACGCTCATGAAGGAGAAGCAGCAGAGCTCGAAGAGCGGCCACAAGAAACAG AAAAACAAGGACAGGGAGTATTACGTGTAAACATGCGAACACTGCTCACACGCGAGTT TTCACAGTTATTTCTATCCACGCCTATGAATCTTTGGACGGTGAGATCTCACAGATGT CAGAACTGCTGGAACCTATGAAATGGGGTATATAACCACGACTCTGGTGGGGAACCG TTTTTTAAAGGACACACACACACACAGCGATGCATCTCTCTCTATAAGCTCAGCCAC GGCTGCGGCAAGGTCCCAGCGGTGCTGGGAGACAGAAGGTTTTGTGCCCTGCTGTAT CATAAAGCACACACTTAGCGCTCTGGAGCCGGA		
	ORF Start: ATG at 61	ORF Stop: TAA at 5074	
	SEQ ID NO: 18	1671 aa	MW at 184075.2 Da
NOV3a, CG108175-01 Protein Sequence	MSSTLHSVFFTLKVSILLGSLGLCLGLEFMGLPNQWARYLRWDASTRSDLSFQFKTN VSTGLLLYLDDGGVCDLFLSLVDGRVQLRFSMDCAETAVALSNKQVNDSSWHFLMVS RDLRLTVLMLDGEQSGELQRPYPMDVVSDFLGGVPTDIRPSALTLDGVPAMPFGK LILDLKYGNSEPRLLGSRGVQMDAEGPCGERPCENGICFLLDGHPTCDCSTTGYGK LCSEGLSHLMSEQAREENVATFRGSEYLCYDLSQNP IQSSSDEITLSFKTWQRNGLI LHTGKSADYVNLAKDGA VSLVINLGS GAFEAI VEPVNGKFNDNAWHDVK VTRNLRQV TISVDGILTTTGYTQEDYTM LGSDDFFVYGGSPSTADLPGSPVSNNFMGCLKEVVYKN NDIRLELSRLARIADTKMKIYGEVVFKEENVATLDPINFETPEAYISLPKWNTKRMGS ISFDFRTTEPNGLILFTHGKPQERKDARSQKNTKVDFFAVELLDGNLYLLDMGSGTI KV KATQKKANDGEWYHVDIQRDGRSGTISVNSRRTPTASGESEILDLEGD MYLGGLP ENRAGLILPTELWTAMLNYGYVGCIRD LFI DGRSKNIRQLAEMQNAAGVKSSCSRMSA KQCDSYPCKNNAVCKDGNRFICDCTGTGYWGRTCEREASILSYDGSMYMKIIMPVM HTEAEDVSFRFMSQRAYGLLVATTSRDSADTLRLLELDGGRVKLMVNLD CIRINCNSSK GPETLYAGQKLNDNEWHTVRVVRGKSLKLTVD DDVAEGTMVGDHTRLEFHN IETGIM TEKRYISVVPSSF IGH LQSLMFNGLLYIDLCKNGDIDYCE LKARFGLRNI IADPVTFR TKSSYLSLATLQAYTSMHLFFQFKTTSPDGFILFNSGDGNDFI AVELVKGYIHYVFDL GNGPNVIKGNSDRPLNDNQWHNVITRDNSNTHSLKVDTKVVTQVINGAKNLDLKGDL YMAGLAQGMYSNLPKLVASRDGFQGC LASVDLNGRLPDLINDALHRSGQIERGCEGT LLGPSTTCQEDSCANQGVCMQWEGFTCDCSMTSYSGNQCNDPGATYIFGKSGGLILY TWPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQGKIGVVFNIGT VDISIKEERTPVNDGKYHVVRFRNNGGNATLQVDNWPVNEHYPTGN TDNERFQMVKQK IPFKYNRPVEEWLQEKGRQLTIFNTQAQIAIGGKDKGR LFPQGQLSGLYDGLKVLNMA AENNPNIKINGSVRLVGEVPSILGTTQTTSMPPEMSTTVMETTTT MATTTTTRKNRSTA SIQPTSDDLVS SAECSSDDEDFVECEPSTGGELVIPLLEDPLATPPIATRAPSITLP PTFRPLLTIIETTKDSLMTSEAGLPCLSDQSGDGDGDDGLVISGYGSGETFD SNLPP TDDDFYTTTFSLVTDKSLSTSIFEGGYKAHAPKWESKDFRPNKVSETSRTTTTSLSPE LIRFTASSSSGMVPKLPAGKMNNRDLKPQPDIVLLPLPTAYELDSTKLKSP LITSPMF RNVPTANPTEPGIRRVPGASEVIRESSSTGMVVGIVAAAALCILILLYAMYKYRNRD EGSYQVDETRNYISNSAQSNGLMKEKQSSKSGHKKQKNKDREYV		
	SEQ ID NO: 19	5335 bp	
NOV3b, CG108175-02 DNA Sequence	CATACAGACAGATCCCAAATCTTCTGTTCAACTGGAAAGGTCTTTTCTCTGGAGTCCT GGGAGGCAAGTTATGGGCAGCACTGCTTCTGGCCGCACCATGAAGCCTGAGTCTGCTT GCGCTCTGCCAGGGCCCTGCTCTGTCTGAGCATTGGGCTTCTAGCTGCCCCCTCCC CACAGCCTGCCGCTGCTAGGAGGTAGAACTTTAGGAGTGGTCCTTGGCCTGTTTCTAC CTGTCACCTGGCTCACCTACCACTCACTCCTCCTCCATCACAGCACCCGGCCCTCC		

CTGTCCCTGGCCTCCCTGGCTGGGGCATTGGGGGTCCGCTGGGAGGAGTGCATCGCT  
 GAAGGCTTCTTCTACTCTCCTGCACCTTCTCCTCCTTGAGTCAAGGCCCTCCGGATCC  
 ACATGGATAGCTGAGATCTTTTCTTGGAGAAAGACGCTTTCCTCTTTACTCCAGTCCC  
 TCACTTCCCCACCTGATTTTCTCCTCTTCTGCTGGTCTGTCTTTTCTACTGCCTC  
 TTTATTCAATTTCTTGCTTGTGTGCCCTCTGGGACTCTCTGTACACTTTCCTCCAT  
 CTCCACTATCTCAGGATCTGTGTGTGTGCTGCCTTCTCCTGTGTGCTTTCTGTCCCC  
 CCATCTCTGTCTTGTCTTTCCCACTTCTATTGCCAAAGGGAGAGATCCTCTCCGGGCT  
 GTTCCCTGGCCTGTCTGCTCCTCCGGGCTCTGTCCCAGCAGCGACAATGAGCTCCACA  
 CTCCACTCGGTTTTCTTACCCTGAAGGTGAGCATCCTGTGGGGTCCCTGCTGGGGC  
 TCTGCCTGGGCCTTGAGTTCATGGGCCTCCCCAACCCAGTGGGCCGCTACCTCCGCTG  
 GGATGCCAGCACACGAGTGACCTGAGTTTCCAGTTCAAGACCAACGCTCTCTACGGGG  
 CTGCTCCTCTACCTGGATGATGGCGGCGTCTGCGACTTCTATGCCTCTCCCTGGTGG  
 ATGGCCGCGTTCAGCTCCGCTTACGATGGACTGTGCGGAGACTGCCGTGCTGTCCAA  
 CAAGCAGGTGAATGACAGCAGCTGGCACTTCTCATGGTGAGCCGTGACCGCCTGCGC  
 ACGGTGCTGATGCTTGATGGCGAGGGCCAGTCTGGGGAGCTGCAGCCCCAGCGGCCCT  
 ACATGGATGTGGTCAGTGACTTGTTCCTTGGTGGAGTCCCTACTGACATACGACCTTC  
 TGCCCTGACCCTTGATGGAGTTTCCAGCCATGCCCGGCTTCAAGGGGTTAATTCTGGAT  
 CTCAAGTATGGAACTCGGAGCCTCGGCTTCTGGGGAGCCGGGTGTCCAGATGGATG  
 CCGAGGGACCCTGTGGTGAGCGTCCCTGTGAAAATGGTGGGATCTGCTTTCTCCTGGA  
 CGGCCACCCACCTGTGACTGTTCACCACTGGCTATGGTGGCAAGCTCTGCTCAGAA  
 GATGTCAGTCAAGATCCAGGCCTCTCCACCTCATGATGAGTGAACAAGGTAGGTGCT  
 TTGCTCGAGAGGAGAATGTGGCCACTTTCCGAGGCTCAGAGTATCTGTGCTACGACCT  
 GTCTCAGAACCCGATCCAGAGCAGCAGTGATGAAATCACCCTCTCCTTTAAGACCTGG  
 CAGCGTAACGGCCTCATCCTGCACACGGGCAAGTCCGCTGACTATGTCAACCTGGCTC  
 TGAAGGATGGTGCGGTCTCCTTGGTTCATTAACCTGGGGTCCGGGGCCTTTGAGGCCAT  
 TGTGGAGCCAGTGAATGGAAAATTCAACGACAACGCCTGGCATGATGTCAAAGTGACA  
 CGCAACCTCCGGCAGGTGACAATCTCTGTGGATGGCATTCTTACCAGCAGGGCTACA  
 CTCAAGAGGACTATACCATGCTGGGCTCGGACGACTTCTTCTATGAGGGAAGCCC  
 AAGTACCGCTGACTTGCTGGCTCCCTGTGAGCAACAACCTTCATGGGCTGCCTTAAA  
 GAGGTTGTTTATAAGAATAATGACATCCGTCTGGAGCTGTCTCGCTGGCCCGGATTG  
 CGGACACCAAGATGAAAATCTATGGCGAAGTTGTGTTAAGTGTGAGAATGTGGCCAC  
 ACTGGACCCCATCACTTTGAGACCCAGAGGCTTACATCAGCTTGCCCAAGTGGAAC  
 ACTAAACGTATGGGCTCCATCTCCTTTGACTTCCGCACCACAGAGCCCAATGGCCTGA  
 TCCTCTTCACTCATGGAAAGCCCCAAGAGAGGAAGGATGCTCGGAGCCAGAAGAATAC  
 AAAAGTAGACTTCTTTGCCGTGGAACCTCTCGATGGCAACCTGTACTTGCTGCTTGAC  
 ATGGGCTCTGGCACCATCAAAGTGAAAGCCACTCAGAAGAAAGCCAATGATGGGGAAT  
 GGTACCATGTGGACATTTCAGCGAGATGGCAGATCAGGTACTATATCAGTGAACAGCAG  
 GCGCACGCCATTACCCGCCAGTGGGGAGAGCGAGATCCTGGACCTGGAAGGAGACATG  
 TACCTGGGAGGGCTGCCGGAGAACCCTGCTGGCCTTATTCTCCCCACCGAGCTGTGGA  
 CTGCCATGCTCAACTATGGCTACGTGGGCTGCATCCGCGACCTATTCAATTGATGGGCG  
 CAGCAAGAACATTCGACAGCTGGCAGAGATGCAGAATGCTGCGGGTGTCAAGTCTCTC  
 TGTTACCGGATGAGTGCCAAGCAGTGTGACAGCTACCCCTGCAAGAATAATGCTGTGT  
 GCAAGGACGGCTGGAACCGCTTCATCTGCGACTGCACCGGCACCGGATACTGGGGAAG  
 AACCTGCGAAAGGGAGGCATCCATCTGAGCTATGATGGTAGCATGTACATGAAGATC  
 ATCATGCCCATGGTCATGCATACTGAGGCAGAGGATGTGTCTTCCGCTTCATGTCCC  
 AGCGAGCTTATGGGCTGCTGGTGGCTACGACCTCCAGGGACTCTGCCGACACCCTGCG  
 TCTGGAGCTGGATGGGGGGCGTGTCAAGCTCATGGTTAACTTAGACTGTATCAGGATA  
 AACTGTAACCTCCAGCAAAGGACCAGAGACCTTGTATGCAGGGCAGAAGCTCAATGACA  
 ACGAGTGGCACACCGTTCCGGGTGGTGGGAGAGGAAAAAGCCTTAAGTTAACCCTGGA  
 TGATGATGTGGCTGAGGGTACAATGGTGGGAGACCATAACCCGTTTGGAGTTCCACAAC  
 ATTGAAACGGGAATCATGACTGAGAAACGCTACATCTCCGTTGTCCCCCTCCAGCTTTA  
 TTGGCCATCTGCAGAGCCTCATGTTTAAATGGCCTTCTCTACATTGACTTGTGCAAAAA  
 TGGTGACATTGATTATTGTGAGCTGAAGGCTCGTTTGGAGTGAAGAACATCATCGCT  
 GACCCTGTACCTTTAAGACCAAGAGCAGCTACCTGAGCCTTGCCACTTTCAGGCTT  
 ACACCTCCATGCACCTCTTCTTCCAGTTCAAGACCACCTCACCAGATGGCTTCATTCT  
 CTTCAATAGTGGTGATGGCAATGACTTCATTGCGAGTGGAGCTTGTCAAGGGGTATATA  
 CACTACGTTTTTGACCTCGGAAACGGTCCCAATGTGATCAAAGGCAACAGTGACCGCC  
 CCTGAATGACAACAGTGGCACAATGTGTCATCACTCGGGACAATAGTAACACTCA  
 TAGCCTGAAAGTGGACACCAAAAGTGGTCACTCAGGTTATCAATGGTGCCAAAAATCTG  
 GATTTGAAAGGTGATCTCTATATGGCTGGTCTGGCCCCAAGGCATGTACAGCAACCTCC  
 CAAAGCTCGTGGCCTCTCGAGATGGCTTTCAGGGCTGTCTAGCATCAGTGGACTTGAA

	TGGACGCCTGCCAGACCTCATCAATGATGCTCTTCATCGGAGCGGACAGATCGAGCGT GGCTGTGAAGGACCCAGTACCACCTGCCAGGAAGATTCATGTGCCAACCAGGGGGTCT GCATGCAACAATGGGAGGGCTTCACCTGTGATTGTTCTATGACCTCTTATTCTGAAA CCAGTGCAATGATCCTGGCGCTACGTACATCTTTGGGAAAAGTGGTGGGCTTATCCTC TACACCTGGCCAGCCAATGACAGGCCAGCACGCGGTCTGACCGCTTGCCGTGGGCT TCAGCACCCTGTGAAGGATGGCATCTTGGTCCGCATCGACAGTGCTCCAGGACTTGG TGACTTCCTCCAGCTTCACATAGAACAGGGGAAAAATTGGAGTTGTCTTCAACATTGGC ACAGTTGACATCTCCATCAAAGAGGAGAGAACCCTGTAAATGACGGCAAATACCATG TGGTACGCTTCACAGGAACGGCGGCAACGCCACCCTGCAGGTGGACAACTGGCCAGT GAATGAACATTATCCTACAGGCAACACTGATAATGAACGCTTCCAAATGGTAAAAACAG AAAATCCCCTTCAAATATAATCGGCCTGTAGAGGAGTGGCTGCAGGAAAAAGGCCGGC AGTTAACCATCTTCAACACTCAGGCGCAAATAGCCATTGGTGGAAAGGACAAAGGACG CCTCTTCCAAGGCCAACTCTCTGGGCTCTATTATGATGGTTTGAAAGTACTGAACATG GCGGCTGAGAACAACCCCAATATTTAAATCAATGGAAGTGTTCGGCTGGTTGGAGAAG TCCCATCAATTTTGGGAACAACACAGACGACCTCCATGCCACCAGAAATGTCTACTAC TGTCATGGAAACCACTACTACAATGGCGACTACCACAACCCGTAAGAATCGCTCTACA GCCAGCATTCAGCCAACATCAGATGATCTTGTTCATCTGCTGAATGTTCAAGTGATG ATGAAGACTTTGTTGAATGTGAGCCGAGTACAGGTAGGTGAGCAAAACCCACGGAGCC GGGAATCAGACGGGTTCCGGGGGCTCAGAGGTGATCCGGGAGTCGAGCAGCACACA GGGATGGTTCGTCGGCATTGTGGCTGCTGCCGCCCTCTGCATCTTGATCCTCCTGTACG CCATGTACAAGTACAGGAACAGGGACGAGGGGTCTATCAAGTGGACGAGACGCGGAA CTACATCAGCAACTCCGCCCAGAGCAACGGCACGCTCATGAAGGAGAAGCAGCAGAGC TCGAAGAGCGGCCACAAGAAACAGAAAAACAAGGACAGGGAGTATTACGTGTAAACAT GCGAACACTGCTCACACGCGAGTTTTTACAGTTATTTCTATCCACGCCCTATGAATCTT TGGACGGTGAGATCTCACAGATGTGAGAACTGCTGGAACATGAAATGGGGTATATAA CCACGACTCTGGTGGGGAAAACCGTTTTTAAAGGACACACACACACACAGCGATG		
	ORF Start: ATG at 743	ORF Stop: TAA at 5156	
	SEQ ID NO: 20	1471 aa	MW at 162660.3 Da
NOV3b, CG108175-02 Protein Sequence	MSSTLHVSFFTLKVSILLGSLGLCLGLEFMGLPNQWARYLRWDASTRSDLSFQFKTN VSTGLLLYLDDGGVCDLFLCLSLVDGRVQLRFMSMDCAETAVLSNQVNDSSWHFLMVS DRLRTVLMLDGEGQSGELQPPQPYMDVVSDFLGGVPTDIRPSALTLDGVQAMPFGK LILDLKYGNSEPRLLGSRGVQMDAEGPCGERPCENGICFLLDGHPTCDCSTTGYGGK LCSEDSQDPGLSHLMMSEQGRCFAREENVATFRGSEYLCYDLSQNP IQSSSDEITLS FKTWQRNGLILHTGKSADYVNLAKDGA VSLVINLGSGAFEAIPEPVNGKFNDNAWHD VKVTRNLRQVTISVDGILTTTGYTQEDYTM LGSDDFFVYVGGSPSTADLPSPVSNNFM GCLKEVVYKNNDIRLELSRLARIADTKMKIYGEVVFKEENVATLDPINFETPEAYISL PKWNTKRMGSISFDFRTTEPNGLILFTHGKPKERKDARSQKNTKVDFFAVELLDGNLY LLDMGSGTIKVKATQKKANDGEWYHVDIQRDGRSGTISVNSRRTFTASGESEILDL EGDMYLGGLPENRAGLILPTELWTAMLNYGYVGCIRDLFIDGRSKNIRQLAEMQNAAG VKSSCSRMSAKQCDSPCKNNVCKDGWNRFCIDCTGTGYWGRTCEREASILSYDGS YMKIIMPMVMHTEAEDVSFRFMSQRAYGLLVATTSRDSADTLRLLELDGGRVKLMVNLD CIRINCNSSKGPETLYAGQKLNDNEWHTVRVVRGKSLKLTVDVDDVAGTVMGDHTRL EFHNIETGIMTEKRYISVVPSSFIGHLQSLMFNGLLYIDLCKNGDIDYCELMKARFLR NIIADPVTFTKSSYLSLATLQAYTSMHLFFQKTTSPDGFILFNSGDGNDFI AVELV KGYIHYVFDLGNPNVIKGNSDRPLNDNQWHNVITRDNSTHSLKVDTKVVTQVING AKNLDLKGDLYMAGLAQGMYSNLPLVASRDGFQGLASVDLNGRLPDLINDALHRSG QIERGCEGPSTTCQEDSCANQGVCMQQWEGFTCDCSMTSYSGNQCNPDGATYIFGKSG GLILYTPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQKGIGVV FNIGTVDISIKEERTPVNDGKYHVVRFRTRNGGNATLQVDNWPVNEHYPTGNTDNERFQ MVKQKIPFKYNRPVEEWLQEKGRQLTIFNTQAQIAIGGKDKGRLFQGLSGLYYDGLK VLNMAAENNPNIKINGSVRLVGEVPSILGTTQTTSMPPMSTTVMETTTTMAATTTTRK NRSTASIQPTSDDLVSSEAECSSDDEDFVECEPSTGRSANPTEPGIRRVPGASEVIRE SSTTGMVVGIVAAAALCILILLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLTLMKE KQSSKSGHKKQKNKDREYV		
	SEQ ID NO: 21	5116 bp	
NOV3c, CG108175-03	CATACAGACAGATCCCAAATCTTCTGTTCAACTGGAAAGGTCTTTTCTCTGGAGTCTT GGGAGGCAAGTTATGGGCAGCACTGCTTCTGGCCGCACCATGAAGCCTGAGTCTGCTT GCGCTCTGCCCAGGGCCCTGCTCTGTCTGAGCATTGGGCTTCTAGCTGCCCCCTCCC		

DNA Sequence	CACAGCCTGCCGCTGCTAGGAGGTAGAACCTTAGGAGTGGTCCTTGGCCTGTTTCTAC CTGTACCTGGCTCACCTCACCACTCACTCCTCCTCCATCACAGCACCCCGGCCCTCC CTGTCCCTGGCCTCCCTGGCTGGGGCATTGGGGGTCCGCTGGGAGGAGTGCATCGCT GAAGGCTTCTTCTACTCTCTGCACCTTCTCCTCCTTGAGTCAAGGCCTCCGGATCC ACATGGATAGCTGAGATCTTTTCTTGGAGAAAGACGCTTTCCTCTTTACTCCAGTCCC TCACTTCCCCACCTGATTTTCTCCTCTTCTGCTGGTCCTGTCTTTTCTACTGCCTC TTTATTCAATTTCTTGCTTGTGTGCCCCCTCTGGGACTCTCTTGTACACTTTCCTCCAT CTCCACTATCTCAGGATCTGTGTGTGTGCTGCCTTCTCCTGTGTGCTTCTGTCCCC CCATCTCTGTCTTGTCTTTCCCACTTCTATTGCCAAAGGGAGAGATCCTCTCCGGGCT GTTCCCTGGCCTGTCTCCTCCGGGCTCTGTCCCAGCAGCACAATGAGCTCCACA CTCCACTCGTTTTCTTACCCTGAAGGTCAGCATCCTGTGTTGGGTCCCTGCTGGGC TCTGCCTGGGCCTTGAGTTCATGGGCCTCCCCAACCAGTGGGCCCGCTACCTCCGCTG GGATGCCAGCACACGCAGTGACCTGAGTTTCCAGTTCAAGACCAACGTCCTACGGGG CTGCTCCTCTACCTGGATGATGGCGGCGTCTGCGACTTCTATGCCTCTCCCTGGTGG ATGGCCGCGTTTCAGCTCCGCTTCAGCATGGACTGTGCCGAGACTGCCGTGCTGTCCAA CAAGCAGGTGAATGACAGCAGCTGGCACTTCTCATGGTGAGCCGTGACCGCCTGCGC ACGGTGCTGATGCTTGATGGCGAGGGCCAGTCTGGGGAGCTGCAGCCCCAGCGGCCCT ACATGGATGTGGTCAGTGACTTGTTCCTTGGTGGAGTCCCTACTGACATACGACCTTC TGCCCTGACCCTTGATGGAGTTCAAGCCATGCCCGGCTTCAAGGGTTAATTCTGGAT CTCAAGTATGGAACTCGGAGCCTCGGCTTCTGGGGAGCCGGGTGTCCAGATGGATG CCGAGGGACCTGTGGTGAGCGTCCCTGTGAAAATGGTGGGATCTGCTTCTCTCTGGA CGGCCACCCACCTGTGACTGTTCTACCACTGGCTATGGTGGCAAGCTCTGCTCAGAA GATGTCAGTCAAGATCCAGGCCTCTCCACCTCATGATGAGTGAACAAGGTAGGTGCT TTGCTCGAGAGGAGAATGTGGCCACTTTCGAGGCTCAGAGTATCTGTGCTACGACCT GTCTCAGAACCCGATCCAGAGCAGCAGTGATGAAATCACCCCTCTCTTTAAGACCTGG CAGCGTAACGGCCTCATCCTGCACACGGGCAAGTCCGCTGACTATGTCAACCTGGCTC TGAAGGATGGTGCGTCTCCTTGGTCATTAACTGGGGTCCGGGGCCTTTGAGGCCAT TGTGGAGCCAGTGAATGGAAAATTCAACGACAACGCTGGCATGATGTCAAAGTGACA CGCAACCTCCGGCAGGTGACAATCTCTGTGGATGGCATTCTTACCAGCAGGGCTACA CTCAAGAGGACTATACCATGCTGGGCTCGGACGACTTCTTCTATGTAGGAGGAAGCCC AAGTACCGCTGACTTGCCTGGCTCCCTGTGAGCAACAACCTTCATGGGCTGCCTTAAA GAGGTGTGTTATAAGAATAATGACATCCGTCTGGAGCTGTCTCGCCTGGCCCGGATTG CGGACACCAAGATGAAAATCTATGGCGAAGTTGTGTTAAGTGTGAGAAATGTGGCCAC ACTGGACCCCATCAACTTTGAGACCCAGAGGCTTACATCAGCTTGCCCAAGTGGAAC ACTAAACGTATGGGCTCCATCTCCTTTGACTTCCGACCCACAGAGCCCAATGGCCTGA TCCTCTTCACTCATGAAAGCCCCAAGAGAGGAAGGATGCTCGGAGCCAGAAGAATAC AAAAGTAGACTTCTTTGCGGTGGAACCTCCTCGATGGCAACCTGTACTTGTGCTTGAC ATGGGCTCTGGCACCATCAAAGTGAAAGCCACTCAGAAGAAAAGCCAATGATGGGGAAT GGTACCATGTGGACATTACAGCGAGATGGCAGATCAGGTACTATATCAGTGAACAGCAG GCGCACGCCATTACCCGCCAGTGGGGAGAGCGAGATCTTGGACCTGGAAGGAGACATG TACCTGGGAGGGGTGCCGAGAAACCGTGTGCTGGCCTTATTCTCCCCACCGAGCTGTGGA CTGCCATGCTCAACTATGGCTACGTGGGCTGCATCCGCGACCTATTATGATGGGCG CAGCAAGAACATTTCGACAGCTGGCAGAGATGCAGAATGCTGCGGGTGTCAAGTCCCTCC TGTTACCGGATGAGTGCCAAGCAGTGTGACAGCTACCCCTGCAAGAATAATGCTGTGT GCAAGGACGGCTGGAACCGCTTCATCTGCGACTGCACCGGCACCGGATACTGGGGAAG AACCTGCGAAAGGGAGGCATCCATCCTGAGCTATGATGGTAGCATGTACATGAAGATC ATCATGCCCCATGGTCATGCATACTGAGGCAGAGGATGTGTCTTCCGCTTCATGTCCC AGCGAGCTTATGGGCTGCTGGTGGCTACGACCTCCAGGGACTCTGCCGACACCCTGCG TCTGGAGCTGGATGGGGGGCGTGTCAAGCTCATGGTTAACTTAGACTGTATCAGGATA AACTGTAACCTCCAGCAAAGGACCAGAGACCTTGTATGCAGGGCAGAAGCTCAATGACA ACGAGTGGCACACCGTTCCGGTGGTGGCGAGAGGAAAAAGCCTTAAGTTAACCGTGGA TGATGATGTGGCTGAGGGTACAATGGTGGGAGACCATAACCCGTTTGGAGTTCCACAAC ATTGAAACGGGAATCATGACTGAGAAACGCTACATCTCCGTTGTCCCCTCCAGCTTTA TTGGCCATCTGCAGAGCCTCATGTTTAAATGGCCTTCTCTACATTGACTGTGTCAAAAA TGGTGACATTGATTATTGTGAGCTGAAGGCTCGTTTTTGGACTGAGGAACATCATCGCT GACCCTGTCACTTTAAGACCAAGAGCAGCTACCTGAGCCTTGCCACTCTTCAGGCTT ACACCTCCATGCACCTCTTCTTCCAGTTCAAGACCACCTCACCAGATGGCTTCATTCT CTTCAATAGTGGTGATGGCAATGACTTCATTGCAGTCGAGCTTGTCAAGGGGTATATA CACTACGTTTTTGACCTCGGAAACGGTCCCAATGTGATCAAAGGCAACAGTGACCGCC CCCTGAATGACAACAGTGGCACAATGTCGTCATCACTCGGGACAATAGTAACACTCA TAGCCTGAAAGTGGACACCAAGTGGTCACTCAGGTTATCAATGGTGCCAAAAATCTG
--------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	GATTTGAAAGGTGATCTCTATATGGCTGGTCTGGCCCAAGGCATGTACAGCAACCTCC CAAAGCTCGTGGCCTCTCGAGATGGCTTTTCAAGGCTGTCTAGCATCAGTGGACTTGAA TGGACGCCTGCCAGACCTCATCAATGATGCTCTTCATCGGAGCGGACAGATCGAGCGT GGCTGTGAAGGACCCAGTACCACCTGCCAGGAAGATTTCATGTGCCAACCAGGGGGTCT GCATGCAACAATGGGAGGGCTTACCTGTGATTGTTCTATGACCTCTTATTCTGGAAA CCAGTGCAATGATCCTGGCGCTACGTACATCTTTGGGAAAAGTGGTGGGCTTATCCTC TACACCTGGCCAGCCAATGACAGGCCAGCACGCGGTCTGACCGCCTTGCCGTGGGCT TCAGACCACCTGTGAAGGATGGCATCTTGGTCCGCACTGCAGACAGTGCTCCAGGACTTGG TGACTTCCCTCCAGCTTACATAGAACAGGGGAAAATTGGAGTTGTCTTCAACATTGGC ACAGTTGACATCTCCATCAAAGAGGAGAGAACCCTGTAAATGACGGCAAATACCATG TGGTACGCTTCACCAGGAACGGCGGCAACGCCACCCTGCAGGTGGACAACCTGGCCAGT GAATGAACATTATCCTACAGGCAACACTGATAATGAACGCTTCCAAATGGTAAAACAG AAAATCCCCTTCAAATATAATCGGCCTGTAGAGGAGTGGCTGCAGGAAAAAGGCCGGC AGTTAACCATCTTCAACACTCAGGCGCAAATAGCCATTGGTGGAAAGGACAAAGGACG CCTCTTCCAAGGCCAACTCTCTGGGCTCTATTATGATGGTTTGAAAGTACTGAACATG GCGGCTGAGAACAACCCCAATATTAAAATCAATGGAAGTGTTCCGCTGGTTGGAGAAG TCCCATCAATTTTGGGAACAACACAGACGACCTCCATGCCACCAGAAATGTCTACTAC TGTCATGGAAACCACTACTACAATGGCGACTACCACAACCCGTAAGAATCGCTCTACA GCCAGCATTCAGCCAACATCAGATGATCTTGTTCATCTGCTGAATGTTCAAGTGATG ATGAAGACTTTGTTGAATGTGAGCCGAGTACAGGTAGGTCAGCCAGAAGCTCTAATGC AGCTAGAATCACTCCGTGCCGCCCTTACATGGACATGGCGACTCACTTACACATTTAC TCCTATCATCTTCATCTCCTGTGTAGTTCACTCATAGATATGACCCTCCCCTTCTCTGC ATCTTTCCTTCCCCATTCTCCCCCTTTCTTTAGCATTGTTAAAATTTATGTGCTGTCA TCCATCTCCC <u>TAAATTAAAGAAAGCCTAAAATTTGTCAAAAAGACAAAAAATATATA</u> <u>TATCTGAAAAC</u>		
	ORF Start: ATG at 743		ORF Stop: TAA at 5057
	SEQ ID NO: 22	1438 aa	MW at 159120.8 Da
NOV3c, CG108175-03 Protein Sequence	MSSTLHSVFFTLKVSILLGSLGLCLGLEFMGLPNQWARYLRWDASTRSDLSFQFKTN VSTGLLLYLDDGGVCDLFLCLSLVDGRVQLRFSMDCAETAVALSNKQVNDSSWHFLMVS RDLRTLVLMLDGEGQSGELQRPYMDVVDLFLGGVPTDIRPSALTLDGVMQAMPFGK LILDLKYGNSEPRLLGSRGVQMDAEGPCGERPCENGIGCFLLDGHPTCDCSTTGYGK LCSEDVSDPGLSHLMSEQGRCFAREENVATFRGSEYLCYDLSQNP IQSSSDEITLS FKTWQRNGLILHTGKSADYVNLALKDGAVALVINLGSFAFEAIVEPVNGKFNDNAWHD VKVTRNLRQVTISVDGILTTTGYTQEDYTMGLSDDFFVVGSPSTADLPSPVSNFNM GCLKEVVYKNNDIRLELSRLARIADTKMKIYGEVVFKEENVATLDPINFETPEAYISL PKWNTKRMGSI SFDFTTEPNGLILFTHGKPKQERKDARSQKNTKVDFFAVELLDGNLY LLDMGSGTIKVKATQKKANDGEWYHVDIQRDGRSGTISVNSRRPTFTASGESEILD EGDMYLGGLPENRAGLILPTELWTAMLNYGVGCI RDLFI DGRSKNIRQEAEMQNAAG VKSSCSRMSAKQCDSYPCKNNAVCKDGWNRFCIDCTGTGYWGRTCEREASILSYDGSM YMKIIMPVMHTEAEDVSFRFMSQRAYGLLVATTSRDSADTLRLELDGGRVKLMVNLD CIRINCNSKGPETLYAGQKLNDNEWHTVRVVRGKSLKLTVDVDDVAEGTMVGDHTRL EFHNIETGIMTEKRYISVVPSSFIGHLQSLMFNGLLYIDLCKNGDIDYCELKARFGLR NIIADPVTFTKSSYLSLATLQAYTSMHLFFQFKTTS PDGFILFNSGDGNDFIAVELV KGYIHYVFDLGNPNVIKGNSDRPLNDNQWHNVITRDNSTHSLKVDTKVVTQVING AKNLDLKGDLYMAGLAQGMYSNLPKLVASRDGFQGCCLASVDLNGRLPD LINDALHRSG QIERGCEGPSTTCQEDSCANQGVCMQWEGFTCDCSMTSYSGNQCNPDGATYIFGKSG GLILYTPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQKGIGVV FNIGTVDISIKEERTPVNDGKYHVVRFTNRNGGNATLQVDNWPVNEHYPTGNTDNERFQ MVKQKIPFKYNRPVEEWLQEKGRQLTIFNTQAQIAIGGKDKGRLFQGGQLSGLYDGLK VLNMAAENPNINIKINGSVRLVGEVPSILGTTQTSMPPPEMSTTVMETTTTMTATTTTRK NRSTASIQPTSDDLVS SAECSSDDEDFVECEPSTGRSARSSNAARITPCRPYMDMATH LHIYSYHLHLCLSSLIDMTLPFLHLSFPILPLSLALLKFMCCHPSP		
	SEQ ID NO: 23		5656 bp
NOV3d, CG108175-04 DNA Sequence	CATACAGACAGATCCCAAATCTTCTGTTCAACTGGAAAGGTCTTTTCTCTGGAGTCCT GGGAGGCAAGTTATGGGCAGCACTGCTTCTGGCCGCACCATGAAGCTGAGTCTGCTT GCGCTCTGCCAGGCGCTGCTCTGTCTGAGCATTGGGCTCTAGCTGCCCCCCCTCCC CACAGCCTGCCGCTGCTAGGAGGTAGAAGTTTAGGAGTGGTCTTGGCCTGTTTCTAC CTGTACCTGGCTCACCTCACTCACTCCTCCTCCATCACAGCACCCCGGCCCTCC		

CTGTCCCTGGCCTCCCTGGCTGGGGCATTGGGGGTCCGCTGGGAGGAGTGCATCGCT  
 GAAGGCTTCTTCTACTCTCCTGCACCTTCTCCTCCTTGAGTCAAGGCCTCCGGATCC  
 ACATGGATAGCTGAGATCTTTTCTTGGAGAAAGACGCTTTCCTCTTTACTCCAGTCCC  
 TCACTTCCCCACCTGATTTTCTCCTCTTCTGCTGGTCTGTCTTTTCTACTGCCTC  
 TTTATTCAATTTCTTGCTTGTGTGCCCCCTCTGGGACTCTCTTGTACACTTTCTCCAT  
 CTCCACTATCTCAGGATCTGTGTGTGTGCTGCCTTCCTCCTGTGTGCTTTCTGTCCCC  
 CCATCTCTGTCTTGTCTTTCCCACTTCTATTGCCAAAGGGAGAGATCCTCTCCGGGCT  
 GTTCCCTGGCCTGTCTGCTCCTCCGGGCTCTGTCCAGCAGCGACAATGAGCTCCACA  
 CTCCACTCGGTTTTCTTACCCTGAAGGTCAGCATCCTGTGGGGTCCCTGCTGTGGGC  
 TCTGCCCTGGGCCTTGAGTTCATGGGCCTCCCCAACAGTGGGCCCTACCTCCGCTG  
 GGATGCCAGCACACGCAGTGACCTGAGTTTCCAGTTCAAGACCAACGTCTCTACGGGG  
 CTGCTCCTCTACCTGGATGATGGCGGCGTCTGCGACTTCTATGCCTCTCCCTGGTGG  
 ATGGCCGCGTTCAGCTCCGCTTACGATGGACTGTGCCGAGACTGCCGTGCTGTCCAA  
 CAAGCAGGTGAATGACAGCAGCTGGCACTTCTCATGGTGAGCCGTGACCGCCTGCGC  
 ACGGTGCTGATGCTTGATGGCGAGGGCCAGTCTGGGGAGCTGCAGCCCCAGCGGCCCT  
 ACATGGATGTGGTCAGTGACTTGTTCCTTGGTGGAGTCCCTACTGACATACGACCTTC  
 TGCCCTGACCCTTGATGGAGTTCAGGCCATGCCCGGCTTCAAGGGGTAAATTCTGGAT  
 CTCAAGTATGGAACTCGGAGCCTCGGCTTCTGGGGAGCCGGGGTGTCCAGATGATG  
 CCGAGGGACCCCTGTGGTGAGCGTCCCTGTGAAAATGGTGGGATCTGCTTTCTCCTGGA  
 CGGCCACCCACCTGTGACTGTTCTACCACTGGCTATGGTGGCAAGCTCTGCTCAGAA  
 GATGTCAGTCAAGATCCAGGCCTCTCCACCTCATGATGAGTGAACAAGGTAGGTGCT  
 TTGCTCGAGAGGAGAATGTGGCCACTTTCGAGGCTCAGAGTATCTGTGCTACGACCT  
 GTCTCAGAACCCGATCCAGAGCAGCAGTGATGAAATCACCTCTCCTTTAAGACCTGG  
 CAGCGTAACGGCCTCATCCTGCACACGGGCAAGTCGGCTGACTATGTCAACCTGGCTC  
 TGAAGGATGGTGCAGTCTCCTTGGTCATTAACCTGGGGTCCGGGGCCTTTGAGGCCAT  
 TGTGGAGCCAGTGAATGGAAAATTCAACGACAACGCCTGGCATGATGTCAAAGTGACA  
 CGCAACCTCCGGCAGGTGACAATCTCTGTGGATGGCATTCTTACCACGACGGGCTACA  
 CTCAAGAGGACTATACCATGCTGGGCTCGGACGACTTCTTCTATGTAGGAGGAAGCCC  
 AAGTACCGCTGACTTGCCTGGCTCCCTGTGAGCAACAACCTTCATGGGCTGCCTTAAA  
 GAGGTTGTTTATAAGAATAATGACATCCGTCTGGAGCTGTCTCGCCTGGCCCGGATTG  
 CGGACACCAAGATGAAAATCTATGGCGAAGTTGTGTTAAGTGTGAGAATGTGGCCAC  
 ACTGGACCCCATCAACTTTGAGACCCCAAGGCTTACATCAGCTTGCCCAAGTGGAAC  
 ACTAAACGTATGGGCTCCATCTCCTTTGACTTCCGCACCACAGAGCCCAATGGCCTGA  
 TCCTCTTCACTCATGGAAAGCCCCAAGAGAGGAAGGATGCTCGGAGCCAGAAGAATAC  
 AAAAGTAGACTTCTTTGCCGTGGAACCTCTCGATGGCAACCTGTACTTGTCTGTGAC  
 ATGGGCTCTGGCACCATCAAAGTGAAAGCCACTCAGAAGAAAGCCTGATGGGGAAT  
 GGTACCATGTGGACATTACGCGAGATGGCAGATCAGGTACTATATCAGTGAACAGCAG  
 GCGCACGCCATTACCGCCAGTGGGGAGAGCGAGATCCTGGACCTGGAAGGAGACATG  
 TACCTGGGAGGGCTGCCGGAGAACCGTGTGTCCTTATTCTCCCCACCGAGCTGTGGA  
 CTGCCATGCTCAACTATGGCTACGTGGGCTGCATCCGCGACCTATTCAATTGATGGGCG  
 CAGCAAGAACATTCGACAGCTGGCAGAGATGCAGAATGCTGCGGGTGTCAAGTCTCTCC  
 TGTTACGGATGAGTGCCAAGCAGTGTGACAGCTACCCCTGCAAGAATAATGCTGTGT  
 GCAAGGACGGCTGGAACCGCTTCATCTGCGACTGCACCGGCACCGGATACTGGGGAAG  
 AACCTGCGAAAGGGAGGCATCCATCCTGAGCTATGATGGTAGCATGTACATGAAGATC  
 ATCATGCCCATGGTCTATGCTACTGAGGCAGAGGATGTGTCCTTCCGCTTCATGTCCC  
 AGCGAGCTTATGGGCTGCTGGTGGCTACGACCTCCAGGGACTCTGCCGACCCCTGCG  
 TCTGGAGCTGGATGGGGGGCGTGTCAAGCTCATGGTTAACTTAGACTGTATCAGGATA  
 AACTGTAACCTCAGCAAAGGACCAGAGACCTTGATGCAGGGCAGAAGCTCAATGACA  
 ACGAGTGGCACACCGTTCCGGTGGTGGGAGAGGAAAAAGCCTTAAGTTAACCGTGGGA  
 TGATGATGTGGCTGAGGGTACAATGGTGGGAGACCATAACCGTTTGGAGTTCCACAAC  
 ATTGAAACGGGAATCATGACTGAGAAACGCTACATCTCCGTTGTCCCTCCAGCTTTA  
 TTGGCCATCTGCAGAGCCTCATGTTTAAATGGCCTTCTCTACATTGACTTGTGCAAAA  
 TGGTGACATTGATTATTGTGAGCTGAAGGCTCGTTTGGACTGAGGAACATCATCGCT  
 GACCCTGTACCTTTAAGACCAAGAGCAGTACCTGAGCCTTGCCACTTTCAGGCTT  
 ACACCTCCATGCACCTCTTCTTCCAGTTCAAGACCACTCACCAGATGGCTTCATTCT  
 CTTCATAGTGGTGTGGAATGACTTCATTGCAGTCGAGCTTGTCAAGGGGTATATA  
 CACTACGTTTTTGACCTCGGAAACGGTCCCAATGTGATCAAAGGCAACAGTGACCGCC  
 CCCTGAATGACAACAGTGGCACAATGTGCTCATCACTCGGGACAATAGTAACACTCA  
 TAGCCTGAAAGTGGACACCAAGTGGTCACTCAGGTTATCAATGGTGCCAAAAATCTG  
 GATTTGAAAGGTGATCTCTATATGGCTGGTCTGGCCCCAAGGCATGTACAGCAACCTCC  
 CAAAGCTCGTGGCCTCTCGAGATGGCTTTCAGGGCTGTCTAGCATCAGTGGACTTGAA

	<p>TGGACGCCTGCCAGACCTCATCAATGATGCTCTTCATCGGAGCGGACAGATCGAGCGT  GGCTGTGAAGGACCCAGTACCACCTGCCAGGAAGATTTCATGTGCCAACCAGGGGGTCT  GCATGCAACAATGGGAGGGCTTCACCTGTGATTGTTCTATGACCTCTTATTCTGAAA  CCAGTGCAATGATCCTGGCGCTACGTACATCTTTGGGAAAAGTGGTGGGCTTATCCTC  TACACCTGGCCAGCCAATGACAGGCCCAGCACGCGGTCTGACCGCTTGCCGTGGGCT  TCAGCACCACCTGTGAAGGATGGCATCTTGGTCCGCATCGACAGTGCTCCAGGACTTGG  TGACTTCCTCCAGCTTCACATAGAACAGGGGAAAATTGGAGTTGTCTTCAACATTGGC  ACAGTTGACATCTCCATCAAAGAGGAGAGAACCCTGTAAATGACGGCAAATACCATG  TGGTACGCTTCACCAGGAACGGCGGCAACGCCACCCTGCAGGTGGACAACTGGCCAGT  GAATGAACATTATCCTACAGGCAACACAGGATAATGAACGCTTCCAAATGGTAAAAACAG  AAAATCCCCCTTCAAATATAATCGGCCTGTAGAGGAGTGGCTGCAGGAAAAAGGCCGGC  AGTTAACCATCTTCAACACTCAGGCGCAAATAGCCATTGGTGGAAAGGACAAAGGACG  CCTCTTCCAAGGCCAACTCTCTGGGCTCTATTATGATGGTTTGAAAGTACTGAACATG  GCGGCTGAGAACAACCCCAATATTAAAATCAATGGAAGTGTTCGGCTGGTTGGAGAAG  TCCCATCAATTTTGGGAACAACACAGACGACCTCCATGCCACCAGAAATGTCTACTAC  TGTCATGGAAACCACTACTACAATGGCGACTACCACAACCCGTAAGAATCGCTCTACA  GCCAGCATTCAGCCAACATCAGATGATCTTGTTCATCTGCTGAATGTTCAAGTGATG  ATGAAGACTTTGTTGAATGTGAGCCGAGTACAGGTAGGTGAGATAAGAGTCTTTCCAC  TTCAATCTTCGAAGGTGGCTACAAAGCACATGCGCCCAAGTGGGAATCCAAGGACTTT  AGACCTAACAAAGTCTCCGAAACTAGTAGGACTACTACCACATCTTTATCCCCCTGAGC  TGATCCGCTTCACAGCTTCCTCCTCGTCTGGGATGGTGGCCAAATTGCCAGCTGGCAA  AATGAATAACCGTGATCTCAAACCCAGCCTGATATAGTCTTGCTTCCGTTGCCCACT  GCCTATGAGCTAGACAGCACCAAACCTGAAGAGCCCACTAATTACTTCCCCCATGTTCC  GTAATGTGCCACAGCAAACCCACGAGCCGGGAATCAGACGGGTCCGGGGGCTC  AGAGGTGATCCGGGAGTCGAGCAGCACAAACAGGGATGGTCTCGGCATTGTGGCTGCT  GCCGCCCTCTGCATCTTGATCCTCCTGTACGCCATGTACAAGTACAGGAACAGGGACG  AGGGGTCTTATCAAGTGGACGAGACGCGGAACACTACATCAGCAACTCCGCCACAGAGCAA  CGGCACGCTCATGAAGGAGAAGCAGCAGAGCTCGAAGAGCGGCCACAAGAAACAGAAA  AACAAAGGACAGGGAGTATTACGTGTAAACATGCGAACACTGCTCACACGCGAGTTTTT  ACAGTTATTTCTATCCACGCTATGAATCTTTGGACGGTGAGATCTCACAGATGTCAG  AACTGCTGGAACATATGAATGGGGTATATAACCACGACTCTGGTGGGGAAAACCGTTT  TTAAAGGACACACACACACACAGCGATG</p>
	<p>ORF Start: ATG at 743      ORF Stop: TAA at 5477</p>
	<p>SEQ ID NO: 24      1578 aa      MW at 174421.6 Da</p>
<p>NOV3d, CG108175-04 Protein Sequence</p>	<p>MSSTLHSVFFTLKVSILLGSLGLCLGLEFMGLPNQWARYLRWDASTRSDLSFQFKTN  VSTGLLLYLDDGGVCDLFLCLSLVDGRVQLRFSMDCAETAVALSNKQVNDSSWHFLMVS  RDLRTVLMLDGEQSGELQRPQPYMDVVSDFLGGVPTDIRPSALTLDGVPQAMPFGK  LILDLKYGNSEPRLLGSRGVQMDAEGPCGERFCENGIGCFLLDGHPCTCDCSTTGYYGK  LCSEDSQDPGLSHLMMSEQGRCFAREENVATFRGSEYLCYDLSQNP IQSSSDEITLS  FKTWQRNGLILHTGKSADYVNLALKDGAVALSVINLGSGAFEAIPEVNGKFNDAWHD  VKVTRNLRQVTISVDGILTTGTGTQEDYTMGLSDDFFVVGSGPSTADLPGSPVSNFM  GCLKEVVYKNNDIRLELSRLARIADTKMKIYGEVVFKECVNATLDPINFETPEAYISL  PKWNTKRMGSISFDFRTTEPNGLILFTHGKPKERKDARSQKNTKVDFFAVELLDGNLY  LLLDMGSGTIKVKATQKKANDGEWYHVDIQRDGRSGTISVNSRRTPTFASGESEILDL  EGDMYLGLPENRAGLILPTELWTAMLNYGYVGCIRDLFIDGRSKNIRQLAEMQNAAG  VKSSCSRMSAKQCDSPCKNNVCKDGNRFICDCTGTGYWGRTCEREASILSYDGSM  YMKIIMPMVMHTEADVSFRFMSQRAYGLLVATTSRDSADTLRLELDGGRVKLMVNLD  CIRINCNSSKGPETLYAGQKLNDNEWHTVRVVRGKSLKLTVDVDDVAEGTMVGDHTRL  EFHNIETGIMTEKRYISVVPSSFIGHLQSLMFNGLLYIDLCKNGDIDYCELFKARFGLR  NIADPVTFTKSSYLSLATLQAYTSMHLFFQFKTSPDGFILFNSGDGNDFIIVELV  KGYIHYVFDLNGPNVIKGNDRPLNDNQWHNVVITRDNSNTHSLKVDTKVVTQVING  AKNLDLKGDLYMAGLAQGMYSNLPKLVASRDGFQGCCLASVDLNGRLPDLINDALHRSG  QIERGCEGPSTTCQEDSCANQGVCMQWEGFTCDCSMTSYSGNQCNPDGATYIFGKSG  GLILYTPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQKIGVV  FNIGTVDISIKEERTPVNDGKYHVVRFTNRNGGNATLQVDNWPVNEHYPTGNTDNERFQ  MVKQKIPFKYNRPVEEWLQEKGRQLTIFNTQAQIAIGGKDKGRFLFQGGQLSGLYYDGLK  VLNMAAENNPNIKINGSVRLVGEVPSILGTTQTTSMPPEMSTTVMETTTTMAATTTTRK  NRSTASIQPTSDDLVSAAECSSDDEDFVECEPSTGRSDKSLSTSIFEGGYKAHAPKWE  SKDFRPKNVSETSRTTTTLSPELIRFTASSSSGMVFKLPAGKMNNDLKPQPDIVLL</p>



	PLPTAYELDSTKLKSPLITSPMFRNVPTANPTEPGIRRVPGASEVIRESSSTTGMVVG IVAAAAALCILILLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLTMKEKQSSKSGH KKQKNKDREYYV	
	SEQ ID NO: 25	4999 bp
NOV3e, CG108175-05 DNA Sequence	CATACAGACAGATCCCAAATCTTCTGTTCAACTGGAAAGGTCTTTTCTCTGGAGTCCT GGGAGGCAAGTTATGGGCAGCACTGCTTCTGGCCGCACCATGAAGCCTGAGTCTGCTT GCGCTCTGCCCAGGGCCCTGCTCTGTCTGAGCATTTGGGCTTCTAGCTGCCCCCTCCC CACAGCCTGCCGCTGCTAGGAGGTAGAAGTTTAGGAGTGGTCTTGGCCTGTTTCTAC CTGTCACTGGCTCACCTCACCACTCACTCCTCCTCCATCACAGACCCCGCCCTCC CTGTCCCTGGCCTCCCTGGCTGGGGCATTGGGGGTCCGCTGGGAGGAGTGCATCGCT GAAGGCTTCTTCTACTCTCCTGCACCTTCTCCTCCTTGAGTCAAGGCTCCGGATCC ACATGGATAGCTGAGATCTTTTCTTGGAGAAAGACGCTTTCCTCTTTACTCCAGTCCC TCATTCCCCACCTGATTTTCTCCTCTTCTGCTGGTCTGTCTTTTCTACTGCCTC TTTATTCAATTTCTTGCTTGTGTGCCCCCTCTGGGACTCTCTTGTAACATTTCTCCAT CTCCACTATCTCAGGATCTGTGTGTGTGCTGCCTTCTCCTGTGTGCTTTCTGTCCCC CCATCTCTGTCTTGTCTTTCCCACTTCTATTGCCAAAGGGAGAGATCCTCTCCGGGCT GTTCCCTGGCCTGTCTGCTCCTCCGGGCTCTGTCCAGCAGCGACAATGAGCTCCACA CTCCACTCGGTTTTCTTACCCTGAAGGTCAGCATCCTGTGGGTCCCTGCTGGGGC TCTGCCTGGGCCTTGAGTTCATGGGCCTCCCCAACCAGTGGGCCCGCTACCTCCGCTG GGATGCCAGCACACGCAGTGACCTGAGTTTCCAGTTCAAGACCAACGTCTCTACGGGG CTGCTCCTCTACCTGGATGATGGCGGCGTCTGCGACTTCCTATGCCTCTCCCTGGTGG ATGGCCGCGTTCAGCTCCGCTTCAGCATGGACTGTGCCGAGACTGCCGTGCTGTCCAA CAAGCAGGTGAATGACAGCAGCTGGCACTTCTCATGGTGAGCCGTGACCGCCTGCGC ACGGTGCTGATGCTTGATGGCGAGGGCCAGTCTGGGGAGCTGCAGCCCCAGCGGCCCT ACATGGATGTGGTCAGTGACTTGTTCCTTGGTGGAGTCCCTACTGACATACGACCTTC TGCCCTGACCCTTGATGGAGTTCAGGCCATGCCCGGCTTCAAGGGGTTAATTTCTGGAT CTCAAGTATGGAACTCGGAGCCTCGGCTTCTGGGGAGCCGGGTGTCCAGATGGATG CCGAGGGACCCTGTGGTGAGCGTCCCTGTGAAAATGGTGGGATCTGCTTTCTCTGGA CGGCCACCCACCTGTGACTGTTCTACCACTGGCTATGGTGGCAAGCTCTGCTCAGAA GATGTCAGTCAAGATCCAGGCCTCTCCCACTCATGATGAGTGAACAAGGTAGGTGCT TTGCTCGAGAGGAGAATGTGGCCACTTTCCGAGGCTCAGAGTATCTGTGCTACGACCT GTCTCAGAACCCGATCCAGAGCAGCAGTGATGAAATCACCCCTCTCCTTTAAGACCTGG CAGCGTAACGGCCTCATCCTGCACACGGGCAAGTCCGGCTGACTATGTCAACCTGGCTC TGAAGGATGGTGCGGTCTCCTTGGTCATTAACCTGGGGTCCGGGGCCTTTGAGGCCAT TGTGGAGCCAGTGAATGGAAATTCAACGACAACCGCTGGCATGTCAAAGTGACA CGCAACCTCCGGCAGGTGACAATCTCTGTGGATGGCATTCTTACCACGACGGGCTACA CTCAAGAGGACTATACCATGCTGGGCTCGGACGACTTCTTCTATGTAGGAGGAAGCCC AAGTACCGCTGACTTGCCCTGGCTCCCTGTGAGCAACAACCTTCATGGGCTGCCTTAAA GAGGTGTTTATAAGAATAATGACATCCGTCTGGAGCTGTCTCGCCTGGCCCGGATTG CGGACACCAAGATGAAAATCTATGGCGAAGTTGTGTTAAGTGTGAGAATGTGGCCAC ACTGGACCCCATCAACTTTGAGACCCCAAGGCTTACATCAGCTTGCCCAAGTGGAAC ACTAAACGTATGGGCTCCATCTCCTTTGACTTCCGCACCACAGAGCCCAATGGCCTGA TCCTCTTCACTCATGGAAAGCCCCAAGAGAGGAAGGATGCTCGGAGCCAGAAGAATAC AAAAGTAGACTTCTTTGCCGTGGAACCTCCTCGATGGCAACCTGTACTTGTCTTGAC ATGGGCTCTGGCACCATCAAAGTGAAAGCCACTCAGAAGAAAGCCCAATGATGGGGAAT GGTACCATGTGGACATTGAGCGAGATGGCAGATCAGGTAATATCAGTGAACAGCAG GCGCACGCCATTCACCGCCAGTGGGGAGAGCGAGATCCTGGACCTGGAAGGAGACATG TACCTGGGAGGGCTGCCGAGAACCGTGTGCTGGCCTTATTCTCCCCACCGAGCTGTGGA CTGCCATGCTCAACTATGGCTACGTGGGCTGCATCCGCGACCTATTCATTGATGGGCG CAGCAAGAACATTCGACAGCTGGCAGAGATGCAGAATGCTGCGGGTGTCAAGTCTCTC TGTTACGGATGAGTGCCAAGCAGTGTGACAGCTACCCCTGCAAGAATAATGCTGTGT GCAAGGACGGCTGGAACCGCTTCATCTGCGACTGCACCGGCACCGGATACGGGGAA AACCTGCGAAAGGGAGGCATCCATCCTGAGCTATGATGGTAGCATGATGAAGATC ATCATGCCCATGGTCATGCATACTGAGGCAGAGGATGTGTCTCCGCTTCATGTCCC AGCGAGCTTATGGGCTGCTGGTGGCTACGACCTCCAGGGACTCTGCCGACCCCTGCG TCTGGAGCTGGATGGGGGGCGTGTCAAGCTCATGGTTAACTTAGACTGTATCAGGATA AACTGTAACCTCAGCAAAGGACCAGAGACCTTGATGCAGGGCAGAAGCTCAATGACA ACGAGTGGCACACCGTTCCGGGTGGTGCAGAGAGGAAAAAGCCTTAAGTTAACCCTGGA TGATGATGTGGCTGAGGGTACAATGGTGGGAGACCATACCCGTTTGGAGTTCCACAAC	



	<p>ATTGAAACGGGAATCATGACTGAGAAAACGCTACATCTCCGTTGTCCCCTCCAGCTTTA  TTGGCCATCTGCAGAGCCTCATGTTTAATGGCCTTCTCTACATTGACTTGTGCAAAAA  TGGTGACATTGATTATTGTGAGCTGAAGGCTCGTTTTGGACTGAGGAACATCATCGCT  GACCCTGTCACCTTTAAGACCAAGAGCAGCTACCTGAGCCTTGCCACTCTTCAGGCTT  ACACCTCCATGCACCTCTTCTCCAGTTCAAGACCACCTCACCAGATGGCTTCATTCT  CTTCAATAGTGGTGATGGCAATGACTTCATTGCAGTCGAGCTTGTCAAGGGGTATATA  CACTACGTTTTTTGACCTCGGAAACGGTCCCAATGTGATCAAAGGCAACAGTGACCGCC  CCCTGAATGACAACAGTGGCACAATGTCGTCATCACTCGGGACAATAGTAACACTCA  TAGCCTGAAAGTGGACACCAAGTGGTCACTCAGGTTATCAATGGTGCCAAAAATCTG  GATTTGAAAGGTGATCTCTATATGGCTGGTCTGGCCCAAGGCATGTACAGCAACCTCC  CAAAGCTCGTGGCCTCTCGAGATGGCTTTCAGGGCTGTCTAGCATCAGTGGAAGTGA  TGGACGCCTGCCAGACCTCATCAATGATGCTCTTCATCGGAGCGGACAGATCGAGCGT  GGCTGTGAAGGACCCAGTACCACCTGCCAGGAAGATTGATGTGCAACAGGGGGTCT  GCATGCAACAATGGGAGGGCTTCACCTGTGATTGTTCTATGACCTCTTATTCTGAAAA  CCAGTGCAATGATCCTGGCGCTACGTACATCTTTGGGAAAAGTGGTGGGCTTATCCTC  TACACCTGGCCAGCCAATGACAGGCCCAGCACGCGGTCTGACCGCCTTGCCGTGGGCT  TCAGCACCACCTGTGAAGGATGGCATCTTGGTCCGCATCGACAGTGTCTCCAGGACTTGG  TGACTTCCTCCAGCTTCACATAGAACAGGGGAAAAATTGGAGTTGTCTTCAACATTGGC  ACAGTTGACATCTCCATCAAAGAGGAGAGAACCCCTGTAAATGACGGCAAATACCATG  TGGTACGCTTCACCAGGAACGGCGGCAACGCCACCCTGCAGGTGGACAACCTGGCCAGT  GAATGAACATTATCCTACAGGCAACACTGATAATGAACGCTTCCAAATGGTAAAAACAG  AAAATCCCTTCAATATAATCGGCCTGTAGAGGAGTGGCTGCAGGAAAAAGGCCGGC  AGTTAACCATCTTCAACACTCAGGCGCAAATAGCCATTGGTGGAAAGGACAAAGGACG  CCTCTTCCAAGGCCAACTCTCTGGGCTCTATTATGATGGTTTGAAAGTACTGAACATG  GCGGCTGAGAACAACCCCAATATTAAATCAATGGAAGTGTTCGGCTGGTTGGAGAAG  TCCCATCAATTTTGGGAACAACACAGACGACCTCCATGCCACCAGAAATGTCTACTAC  TGTCATGGAACCACTACTACAATGGCGACTACCACAACCCGTAAGAATCGCTCTACA  GCCAGCATTGAGCAACATCAGATGATCTTGTTCATCTGCTGAATGTTCAAGTGATG  ATGAAGACTTTGTTGAATGTGAGCCGAGTACAGGTAGGTGAGTAAGAAATGACAACAA  AAAAAGCAAGTTACAAGAATGTGGCAATTCTATTTGTCCAAGAGCATTCTTACACAAC  TTTCTTTTGTAAATTTTTCTTTCATGCCAAAAACATGCGGGCAATTTGTTGATGTAA  GTTGACTATAA</p>
	<p>ORF Start: ATG at 743      ORF Stop: TAA at 4940</p>
	<p>SEQ ID NO: 26      1399 aa      MW at 154757.5 Da</p>
NOV3e, CG108175-05 Protein Sequence	<p>MSSTLHVSFFTLKVSILLGSLGLCLGLEFMGLPNQWARYLRWDASTRSDLSFQFKTN  VSTGLLLYLDGQVDFLCLSLVDGRVQLRFSMDCAETAVALSNKQVNDSSWHFLMVSR  DRLRTVLMLDGEGQSGELQRPQPYMDVVSDFLGGVPTDIRPSALTLDGTVQAMPGFKG  LILDLKYGNSEPRLLGSRGVQMDAEGPCGERPCENGIGCFLLDGHPCTDCSTTGYGGK  LCSEDVSDPGLSHLMMSEQGRCFAREENVATFRGSEYLCYDLSQNP IQSSSDEITLS  FKTWQRNGLILHTGKSADYVNLALKDGAVALVINLGSFAFEAIEPVNGKFNDNAWHD  VKVTRNLRQVTISVDGILTTTGYTQEDYTMGLGSDDFVYVGGSPSTADLPGSPVSNFM  GCLKEVVYKNNDIRLELSRLARIADTKMKIYGEVVFKEENVATLDPINFETPEAYISL  PKWNTKRMGSISFDFRTTEPNGLILFTHGKPKERKDARSQKNTKVDFFAVELLDGNLY  LLDMGSGTIKVKATQKKANDGEWYHVDIQDGRSGTISVNSRPTFTASGESEILD  EGDMYLGGLPENRAGLILPTTELWTAMLNYGYVGCIRDLEFDGRSKNIRQLAEMQNAAG  VKSSCSRMSAKQCDSYPCKNNVCKDGNRFFICDCTGTGYWGRTCEREASILSYDGSM  YMKIIMPMVMHTEAEDVSFRFMSQRAYGLLVATTSRDSADTLRLLELDGGRVKLMVNL  CIRINCNSSKGPETLYAGQKLNDNEWHTVRVVRGKSLKLTVDVDDVAEGTMVGDHTRL  EFHNIEGTIMTEKRYISVVPSSFIGHLQSLMFNGLLYIDLCKNGDIDYCEKARFGLR  NIADPVTFTKSSYLSLATLQAYTSMHLFFQFKTSPDGFILFNSGDGNDFAVELV  KGYIHYVFDLGNPNVIKNSDRPLNDNQWHNVITRDNSTHSLKVDTKVVTQVING  AKNLDLKGDLYMAGLAQGMYSNPKLVASRDGFQGCCLASVDLNGRLPDLINDALHRSG  QIERGCEGPSTTCQEDSCANQGVCMQWEGFTCDCSMTSYSGNQCNPDGATYIFGKSG  GLILYTPANDRPSTRSDRLAVGFSTTVKDGILVRIDSAPGLGDFLQLHIEQGKIGVV  FNIGTVDISIKEERTPVNDGKYHVVRFTNRNGGNATLQVDNWPVNEHYPTGNTDNERFQ  MVKQKIPFKYNRPVEEWLQEKGRQLTIFNTQAQIAIGGKDKGRLFQGGQLSGLYYDGLK  VLNMAAENNPNIKINGSVRLVGEVPSILGTTQTTSMPPEMSTTVMETTTTMAATTTTRK  NRSTASIQPTSDDLVSSAECSSDDEDFVECEPSTGRSVRNDNPKSKLQCEGSPNIPRA  FLHNFL</p>

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

<b>Table 3B. Comparison of NOV3a against NOV3b through NOV3e.</b>		
<b>Protein Sequence</b>	<b>NOV3a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV3b	1..1364 1..1369	1315/1374 (95%) 1315/1374 (95%)
NOV3c	1..1364 1..1369	1315/1374 (95%) 1315/1374 (95%)
NOV3d	1..1364 1..1369	1315/1374 (95%) 1315/1374 (95%)
NOV3e	1..1364 1..1369	1315/1374 (95%) 1315/1374 (95%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

<b>Table 3C. Protein Sequence Properties NOV3a</b>	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 28 and 29

5 A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

<b>Table 3D. Geneseq Results for NOV3a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV3a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE17600	Human extracellular messenger (XMES)-2 protein - Homo sapiens, 1438 aa. [WO200194587-A2, 13-DEC-2001]	1..1363 1..1338	1328/1373 (96%) 1328/1373 (96%)	0.0
AAU28190	Novel human secretory protein, Seq ID No 359 - Homo sapiens, 1712 aa. [WO200166689-A2, 13-SEP-2001]	16..1671 17..1712	1093/1724 (63%) 1324/1724 (76%)	0.0
AAU14241	Human novel protein #112 - Homo sapiens, 1091 aa. [WO200155437-A2, 02-AUG-2001]	368..1363 1..991	990/996 (99%) 990/996 (99%)	0.0

AAU14240	Human novel protein #111 - Homo sapiens, 1061 aa. [WO200155437-A2, 02-AUG-2001]	368..1363 1..961	960/996 (96%) 960/996 (96%)	0.0
AAM79855	Human protein SEQ ID NO 3501 - Homo sapiens, 1522 aa. [WO200157190-A2, 09-AUG-2001]	16..1365 65..1419	952/1392 (68%) 1108/1392 (79%)	0.0

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

**Table 3E. Public BLASTP Results for NOV3a**

Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
A48216	neurexin III-alpha secreted type 1 precursor - rat, 1438 aa.	1..1364 1..1369	1333/1374 (97%) 1346/1374 (97%)	0.0
B48218	neurexin III-alpha membrane-bound type 3 precursor - rat, 1471 aa.	1..1364 1..1369	1333/1374 (97%) 1346/1374 (97%)	0.0
I48216	neurexin III-alpha membrane-bound type 1 precursor - rat, 1578 aa.	1..1364 1..1369	1333/1374 (97%) 1346/1374 (97%)	0.0
Q9Y4C0	Neurexin 3-alpha precursor (Neurexin III-alpha) - Homo sapiens (Human), 1541 aa.	1..1367 1..1338	1328/1373 (96%) 1329/1373 (96%)	0.0
Q07310	Neurexin 3-alpha precursor (Neurexin III-alpha) - Rattus norvegicus (Rat), 1578 aa.	1..1364 1..1369	1318/1374 (95%) 1334/1374 (96%)	0.0

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

**Table 3F. Domain Analysis of NOV3a**

Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_G	55..174	37/152 (24%) 80/152 (53%)	1.5e-11
EGF	202..234	15/47 (32%) 22/47 (47%)	0.0033
laminin_G	281..410	41/161 (25%) 92/161 (57%)	2.5e-22
laminin_G	469..616	53/169 (31%) 112/169 (66%)	2.5e-30

EGF	641..673	10/47 (21%) 26/47 (55%)	0.016
laminin_G	730..840	31/137 (23%) 86/137 (63%)	2.1e-05
laminin_G	893..1024	49/164 (30%) 104/164 (63%)	1.2e-18
EGF	1052..1084	13/47 (28%) 25/47 (53%)	0.0034
laminin_G	1121..1196	26/89 (29%) 52/89 (58%)	1e-06

**Example 4.**

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis		
	SEQ ID NO: 27	2681 bp
NOV4a, CG108624-01 DNA Sequence	CTGGGATGTACCTTTCATCTGTTGCTGCTTTCTTCTATGGGCCCCCTGCCCTCACTCT CAAGAACCTCAACTACTCCGTGCCGAGGAGCAAGGGCCGGCAGGTGATCGGGAAC ATCGGCAGGGATGCTCGACTGCAGCCTGGGCTTCCGCCTGCAGAGCGCGGCGGGAG GGCGCAGCAAGTCGGGTAGCTACCGGGTGCTGGAGAACTCCGCACCGCACCTGCTGGA CGTGGACGCAGACAGCGGGCTCCTCTACACCAAGCAGCGCATCGACCGCGAGTCCCTG TGCCGCCACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTTCCGCCAACGACAAGGAGA TCTGCATGATCAAGGTAGAGATCCAGGACATCAACGACAACGCGCCCTCCTTCTCCTC GGACCAGATCGAAATGGACATCTCGGAGAACGCTGCTCCGGCACCCGCTTCCCCCTC ACCAGCGCACATGACCCGACGCCGGCGAGAATGGGCTCCGCACCTACCTGCTCACGC GCGACGATCACGGCCTCTTTGGAAGTGGACGTTAAGTCCCGCGGCGACGGCACCAAGTT CCCAGAACTGGTCATCCAGAAGGCTCTGGACCGGAGCAACAGAATCACCATACGCTC GTGCTGACTGCCCTGGACGGTGGCGAGCCTCCACGTTCCGCCACCGTACAGATCAACG TGAAGGTGATTGACTCCAACGACAACAGCCCCGGTCTTCGAGGCGCCATCCTACTTGGT GGAAGTGGCCGAGAACGCTCCGCTGGGTACAGTGGTCATCGATCTGAACGCCACCGAC GCCGATGAAGGTCCCAATGGTGAAGTGCTCTACTCTTTCAGCAGCTACGTGCCCTGACC GCGTGCGGGAGCTCTTCTCCATCGACCCCAAGACCGGCCTAATCCGTGTGAAGGGCAA TCTGGACTATGAGGAAAACGGGATGCTGGAGATTGACGTGCGAGCCGAGACCTGGGG CCTAACCCCTATCCAGCCCACTGCAAAGTCACGGTCAAGCTCATCGACCGCAACGACA ATGCGCCGTCCATCGTTTCGTCTCCGTGCGCCAGGGGGCGCTGAGCGAGGCCGCCCC TCCCGGCACCGTCATCGCCCTGGTGCGGGTCACTGACCGGGACTCTGGCAAGAACGGA CAGCTGCAGTGTGCGGTCTTAGCGGAGGAGGGACGGGCGGCGGGGGCCCTGGGCG GGCCCGGGGGTTCCGTCCCCTTCAAGCTTGAGGAGAACTACGACAACCTTCTACACGGT GGTGACTGACCGCCCGCTGGACCGCGAGACACAAGACGAGTACAACGTGACCATCGTG GCGCGGGACGGGGGCTCTCCTCCCCTCAACTCCACCAAGTCGTTCCGGATCAAGATTC TAGACGAGAACGACAACCCGCCTCGGTTACCAAAGGGCTCTACGTGCTTCAGGTGCA CGAGAACAACATCCCGGGAGAGTACCTGGGCTCTGTGCTCGCCAGGATCCCGACCTG GGCCAGAACGGCACCGTATCCTACTCTATCCTGCCCTCGCACATCGGCGACGTGTCTA TCTACACCTATGTGTCTGTGAATCCACGAACGGGGCCATCTACGCCCTGCGCTCCTT TAACCTCGAGCAGACCAAGGCTTTTGAAGTCAAGGTGCTTGCTAAGGACTCGGGGGCG CCCGCGCACTTGGAGAGCAACGCCACGGTGAGGGTGACAGTGCTAGACGTGAATGACA ACGCGCCAGTGATCGTGCTCCCCACGCTGCAGAACGACACCGCGGAGCTGCAGGTGCC GCGCAACGCTGGCCTGGGCTATCTGGTGAGCACTGTGCGCGCCCTAGACAGCGACTTC GGCGAGAGCGGGCGTCTACCTACGAGATCGTGACGGCAACGACGACCACCTGTTTG AGATCGACCCGTCCAGCGGCGAGATCCGACGCTGCACCCCTTCTGGGAGGACGTGAC GCCCGTGGTGGAGCTGGTGGTGAAGGTGACCGACCGCAAGCCTACCCTGTCCGCA	

	GTGGCCAAGCTCATCATCCGCTCGGTGAGCGGATCCCTTCCCAGGGGGTACCACGGG TGAATGGCGAGCAGCACCCTGGGACATGTCTGCTGCCGCTCATCGTGACTCTGAGCAC TATCTCCATCATCCTCTAGCGGCCATGATCACCATCGCCGTCAAGTGCAAGCGCGAG AACAAGGAGATCCGCACTTACAACCTGCCGCATCGCCGAGTACAGCCACCCGAGCTGG GTGGGGGCAAGGGCAAGAAGAAGAAGATCAACAAAAATGATATCATGCTGGTGCAGAG CGAAGTGGAGGAGAGGAACGCCATGAACGTCATGAACGTGGTGAGCAGCCCCCTCCCTG GCCACCTCCCCCATGTACTTCGACTACCAGACCCGCTGCCCTCAGCTCGCCCCGGT CGGAGGTGATGTATCTCAAACCGGCCTCCAACAACCTGACTGTCCCTCAGGGGCACGC GGGCTGCCACACCAGCTTCACCGGACAAGGGACTAATGCAAGCGAGACCCCTGCCACT CGGATGTCCATAATTCAGACAGACAATTTCCCGCAGAGCCCAATTACATGGGCAGCA GGCAGCAGTTTGTTCATGTATTTTCAGTAGCTCCACGTTTAAGGACCCAGAAAGAGCC AGCCTGAGAGACA		
	ORF Start: ATG at 6	ORF Stop: TGA at 2673	
	SEQ ID NO: 28	889 aa	MW at 96584.6 Da
NOV4a, CG108624-01 Protein Sequence	MYLSICCCFLLWAPALTLKLNLSVP EEGAGTVIGNIGRDARLQPLPPAERGGGGR SKSGSYRVLENSAPHLLDVDADSGLLYTKQRIDRESLCRHNACQLSLEVFANDKEIC MIKVEIQDINDNAPSFSSDQIEMDISENAAPGTRFPLTSAHDPDAGENGLRITYLLTRD DHGLFGLDVKSRGDGTFPELVIQKALDREQNHHTLVLTALDGGEP PRSATVQINVK VIDSNDNSPVFEAPSYLVELPENAPLGTVVIDLNATDADEGPNGEVLYSFSSYPDRV RELFSIDPKTGLIRVKGNDLDEENGMLEIDVQARDLGNPIPAHCKVTVKLIDRNDNA PSIGFVSVRQGALSEAAPP GTVIALVRVTD RDSGKNGQLQCRVLGGGGTGGGGGLGGP GGSVPFKLEENYDNFYTVVTD RPLDRETQDEYNVTIVARDGGS PPLNSTKSFAIKILD ENDNPPRFTKGLYVLQVHENNIPGEYLGSVLAQDPDLGQNGTVSYSLPSHIGDVS IY TYVSVNPTNGAIYALRSFNFEQTKAFEFKVLAKDSGAPAHLESNATVRVTVLDVNDNA PVIVLPTLQNDTAELQVPRNAGLGYLVSTVRALDSDFGESGR LTYEIVDGNDDHLFEI DPSSGEIRTLHPFWEDVTPVVELVVKVTDHGKPTLSAVAKLIIRSVSGSLPEGVPRVN GEQHHWDMSLPLIVTLSTISII LLAAMITIAVKCKRENKEIRTYNCRIA EYSHPQLGG GKGKKKKINKNDIMLVQSEVEERNAMNVMNVSSPSLATS PMYFDYQTRLPLSSPRSE VMYLKPASNNLTV PQGHAGCHTSFTGQGTNASETPATRMSIQTDNFPAEPNYMGSRQ QFVQCISVAPRLRTQKEPA		

Further analysis of the NOV4a protein yielded the following properties shown in Table 4B.

Table 4B. Protein Sequence Properties NOV4a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4C.

Table 4C. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY21687	Cadherin-like polypeptide, ontherin -	1..889	880/889 (98%)	0.0

	Vertebrata, 889 aa. [WO9929853-A1, 17-JUN-1999]	1..889	885/889 (98%)	
AAY24913	Human ontherin - Homo sapiens, 889 aa. [WO9929860-A1, 17-JUN-1999]	1..889 1..889	880/889 (98%) 885/889 (98%)	0.0
AAE17313	Human protocadherin protein, sbg419582PROTOCOLADHERIN #2 - Homo sapiens, 855 aa. [WO200198342-A1, 27-DEC-2001]	10..874 14..844	466/869 (53%) 600/869 (68%)	0.0
AAE17312	Human protocadherin protein, sbg419582PROTOCOLADHERIN #1 - Homo sapiens, 888 aa. [WO200198342-A1, 27-DEC-2001]	10..840 14..857	460/882 (52%) 584/882 (66%)	0.0
AAU19545	Human diagnostic and therapeutic polypeptide (DITHP) #131 - Homo sapiens, 427 aa. [WO200162927-A2, 30-AUG-2001]	499..889 36..427	370/392 (94%) 373/392 (94%)	0.0

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4D.

**Table 4D. Public BLASTP Results for NOV4a**

Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O14917	Protocadherin 68 - Homo sapiens (Human), 889 aa.	1..889 1..889	880/889 (98%) 883/889 (98%)	0.0
Q8TAB3	BA99E24.1.1 (Protocadherin 19 (KIAA1313) protein) - Homo sapiens (Human), 1094 aa (fragment).	10..877 7..840	467/872 (53%) 601/872 (68%)	0.0
Q9P2E7	KIAA1400 protein - Homo sapiens (Human), 1093 aa (fragment).	10..873 62..948	394/918 (42%) 558/918 (59%)	0.0
Q96SF0	Protocadherin 10 - Homo sapiens (Human), 896 aa.	10..838 9..859	385/881 (43%) 541/881 (60%)	0.0
Q92518	OL-protocadherin isoform - Mus musculus (Mouse), 1040 aa.	10..873 9..895	393/918 (42%) 553/918 (59%)	0.0

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4E.

**Table 4E. Domain Analysis of NOV4a**

Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
-------------	--------------------	-------------------------------------------------	--------------

cadherin	137..234	30/111 (27%) 74/111 (67%)	2.3e-17
cadherin	248..342	41/110 (37%) 72/110 (65%)	5.4e-22
cadherin	357..463	37/119 (31%) 86/119 (72%)	1.3e-16
cadherin	477..574	33/112 (29%) 71/112 (63%)	1.2e-13
cadherin	593..685	38/108 (35%) 64/108 (59%)	1.3e-10

**Example 5.**

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 29	718 bp	
NOV5a, CG108771-01 DNA Sequence	AAAACTAAGCCTGCTTCCAGTCCCCNCGGGAGTCGTAGGAACCCGTTCTCTGGACGCT GACGTCGGCTTTTCAGGGATCCCTCGCCGGACGCCGCGGAGGGACAGAGCCTGGGAAGC CGTCGCCCCGCCCCGTCCCCGCCCCCGCGCGCAGCGGGCCCCGGGGCGCTGAGACCCGC GTAGAGCAAAGCGCAAGGTCCCAGCGCCCCCTTGGATCCTCGTGGCAGGGTCCGGGCA AGTGTCATTGCGAGGGTTTCAGGAAGCCCCGGCCTGTGATCGTGAGCGGAAACCCCTCC TGGAGTTTCCCCAAAGCCATGGACAGCCCTAGTCTTCGTGAGCTTCAACAGCCTCTGC TGGAGGGCACAGAATGTGAGACCCCTGCCCAGAAGCCTGGCAGGCATGAGCTGGGGTC CCCCCTTAAGAGAGATAGCCTTTGCCGAGTCCCTGAGGGGTTTGCAGTTCCTGTACCCG CCTCTTCCCTCCGTGAGCGCTGGCCTGGGGGAACCAAGGCCCTGATGTTGAGGACA TGTCATCCAGTGACAGTGACTCGGACTGGGATGGAGGCAGCCGCTCTTTCACCATTTCT ACCCACGACCACCTCGGCTTGGCTGTCTTCTCCATGCTGTGTGTTGTTTCTGGCCCGTT GGCATCGCTGCCTTCTGTCTAGCCCAGAAGGTCAGTCTGTGTGTGGGACTTGGAGGGG ACTGGAAGCAGGCTTAGTTTTT		
	ORF Start: ATG at 309	ORF Stop: TAG at 711	
	SEQ ID NO: 30	134 aa	MW at 14376.1 Da
NOV5a, CG108771-01 Protein Sequence	MDSPSLRELQQPLLEGTECETPAQKPGRHELGSPLREIAFAESLRGLQFLSPPLPSVS AGLGEPRPPDVEDMSSSDSDSDWDGGSRLSPFLPHDHLGLAVFSMLCCFWPVGIAAFC LAQKVS LCVGLGGDWKQA		

Further analysis of the NOV5a protein yielded the following properties shown in

## 5 Table 5B.

Table 5B. Protein Sequence Properties NOV5a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.4412 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5C.

<b>Table 5C. Geneseq Results for NOV5a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV5a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB90246	Human polypeptide SEQ ID NO 2622 - Homo sapiens, 172 aa. [WO200190304-A2, 29-NOV-2001]	1..122 1..122	120/122 (98%) 121/122 (98%)	1e-67
AAB25755	Human secreted protein sequence encoded by gene 33 SEQ ID NO:144 - Homo sapiens, 172 aa. [WO200043495-A2, 27-JUL-2000]	1..122 1..122	120/122 (98%) 121/122 (98%)	1e-67
AAB25754	Human secreted protein sequence encoded by gene 33 SEQ ID NO:143 - Homo sapiens, 57 aa. [WO200043495-A2, 27-JUL-2000]	15..71 1..57	57/57 (100%) 57/57 (100%)	2e-27
AAB25697	Human secreted protein sequence encoded by gene 33 SEQ ID NO:86 - Homo sapiens, 101 aa. [WO200043495-A2, 27-JUL-2000]	72..122 1..51	49/51 (96%) 50/51 (97%)	3e-24
AAB43155	Human ORFX ORF2919 polypeptide sequence SEQ ID NO:5838 - Homo sapiens, 88 aa. [WO200058473-A2, 05-OCT-2000]	86..122 2..38	35/37 (94%) 36/37 (96%)	2e-15

- In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5D.

<b>Table 5D. Public BLASTP Results for NOV5a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV5a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9H7V2	CDNA FLJ14220 fis, clone NT2RP3003828 - Homo sapiens (Human), 258 aa.	75..128 161..214	28/54 (51%) 34/54 (62%)	7e-09
Q9H514	BA526K17.1 (Novel protein) - Homo sapiens (Human), 206 aa (fragment).	75..120 161..206	26/46 (56%) 31/46 (66%)	5e-08
O35449	Hypothetical 31.4 kDa protein - Mus musculus (Mouse), 306 aa.	92..128 220..256	16/37 (43%) 23/37 (61%)	0.005



Q96NQ8	CDNA FLJ30323 fis, clone BRACE2007109, highly similar to extensin-like protein NG5 - Homo sapiens (Human), 306 aa.	92..128 220..256	16/37 (43%) 23/37 (61%)	0.005
Q96DW3	Similar to chromosome 6 open reading frame 31 - Homo sapiens (Human), 225 aa.	92..128 139..175	16/37 (43%) 23/37 (61%)	0.005

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

#### Example 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis			
	SEQ ID NO: 31	1174 bp	
NOV6a, CG108782-01 DNA Sequence	ACGCGTGGGCGGACGCGTGGTTGGACTCCGCCCCGTGGAGCCCTGGGCCTGTTGACCCA CCAGCTTAGGAGCACCCACCAAGCTCTGGGTCAACGTGGAGGTACCAGGCCACCATGC TCAGTCTCAAGCTGCCCAACTTCTTCAAGTCCACCAGGTCCCCGGGTGTTCTGGGA AGATGGCATCATGTCTGGCTACCGCCGCCCCACCAGCTCGGCTTTGGACTGTGTCCCTC AGCTCCTTCCAGATGACCAACGAGACGGTCAACATCTGGACTCACTTCCTGCCACCT GGTACTTCTGTGGCGGCTCCTGGCGCTGGCGGGCGGCCCGGCTTCCGTGCGGAGCC GTACCACTGGCCGCTGCTGGTCTTCTGTCTGCCCGCTGCCTCTACCCCTTCGCGTCG TGCTGCGCGCACACCTTCAGCTCCATGTGCCCCGATGCGCCACATCTGCTACTTCC TCGACTACGGCGCGCTCAGCCTCTACAGTCTGGGCTGCGCCTTCCCTATGCCGCTA CTCCATGCCGGCCTCCTGGCTGCACGGCCACCTGCACCAGTTCTTTGTGCTGCGGCC GCACTCAACTCCTTCTGTGCACCGGCTCTCCTGCTACTCCCGGTTCTTGAGCTGG AAAGCCCTGGGCTCAGTAAGGTCTCCGCACAGGAGCCTTCGCCTATCCATTCCTGTT CGACAACCTCCCACTCTTTTATCGGCTCGGGCTGTGCTGGGGCAGGGGCCACGGCTGT GGGCAGGAGGCCCTGAGCACCAGCCATGGCTACCATCTCTTCTGCGCGCTGCTCACTG GCTTCCTCTTTCGCTCCACCTGCCTGAAAGGCTGGCACCAGGACGCTTTGATTACAT CGGTACAGCCACCAGTTATTCCACATCTGTGCAGTGCTGGGCACCCACTTCCAGCTG GAGGCAGTGCTGGCTGATATGGGATCACGCAGAGCCTGGCTGGCCACACAGGAACCTG CCCTGGGCCTGGCAGGCACAGTGGCCACACTGGTCTTGGCTGCAGCTGGGAACCTACT CATTATTGCTGCTTTACAGCCACCCTGCTTCGGGCCCCCAGTACATGCCCTCTGCTG CAGGGTGGCCCACTGGAGGGGGGTACCCAGGCCAAACAACAGTGAGGCCCATCCCTG ACCTGTCTCTGGAG		
	ORF Start: ATG at 113	ORF Stop: TGA at 1145	
	SEQ ID NO: 32	344 aa	MW at 37988.7 Da
NOV6a, CG108782-01 Protein Sequence	MLSLKLPQLLQVHVPRVFWEDGIMS̃GYRRPTSSALDCVLSSFQMTNETVNIWTHFLP TWYFLWRLALAGGPGFRAEPYHWPLLVLFPACLYPFASCCAHTFSSMSPRMRHICY FLDYGALSLSLGAFFPYAAYSMPASWLHGLHQLQFFVPAALNSFLCTGLSCYSRFILE LESPGLSKVLRGTAFAYPFLFDNLPLFYRLGLCWGRGHGCGQEALSTSHGYHLFCALL		

	TGFLFASHLPERLAPGRFDYIGHSHQLFHHICAVLGTGHFQLEAVLADMGSRRRAWLATQE PALGLAGTVATLVLAAGNLLIIAAFTATLLRAPSTCPLLQGGPLEGGTQAKQQ		
	SEQ ID NO: 33	1081 bp	
NOV6b, CG108782-02 DNA Sequence	CAAGCTCTGGGTCAACGTGGAGGTACCAGGCCACCATGCTCAGTCTCAAGCTGCCCCA ACTTCTTCAAGTCCACCAGGTCCCCGGGTGTTCTGGGAAGATGGCATCATGTCTGGC TACCGCCGCCCCACCAGCTCGGCTTTGGACTGTGTCTCAGCTCCTTCCAGATGACCA ACGAGACGGTCAACATCTGGACTCACTTCCTGCCCACCTGGTACTTCTGTGGCGGCT TCTGGCGCTGGCGGGCGGCCCGGCTTCCGTGCGGAGCCGTACCACTGGCCGCTGCTG GTCTTCCTGCTGCCCGCCTGCCTCTACCCCTTCGCGTCGTGCTGCGCGCACACCTTCA GCTCCATGTGCGCCGCGCATGCGCCACATCTGCTACTTCTCGACTACGGCGCGCTCAG CCTCTACAGTCTGGGCTGCGCCTTCCCCTATGCCGCTACTCCATGCCGGCCTCCTGG CTGCACGGCCACCTGCACCAGTTCTTTGTGCCTGCCGCCGCACTCAACTCCTTCCTGT GCACCGGCCTCTCCTGCTACTCCCGTTTCTGGAGCTGGAAAGCCCTGGGCTCAGTAA GGTCTCTCCGCACAGGAGCCTTCGCCTATCCATTCTGTTCGACAACCTCCCACTCTTT TATCGGCTCGGGCTGTGCTGGGGCAGGGGCCACGGCTGTGGGCAGGAGGCCCTGAGCA CCAGCCATGGCTACCATCTCTTCTGCGCGCTGCTCACTGGCTTCCTCTTCGCCTCCCA CCTGCCTGAAAGGCTGGCACCAGGACGCTTTGATTACATCGGCCACAGCCACCAGTTA TTCCACATCTGTGCAGTGTGGGCACCCACTTCCAGCTGGAGGCAGTGTGGCTGATA TGGGATCACGCAGAGCCTGGCTGGCCACACAGGAACCTGCCCTGGGCCTGGCAGGCAC AGTGGCCACACTGGTCTTGGCTGCAGCTGGGAACCTACTCATTATTGCTGCTTTCACA GCCACCCTGCTTCGGGCCCCCGGTACATGCCCTCTGCTGCAGGTGGCCCACTGGAGG GGGTACCCAGGCCAAACAACAGTGAGGCCCCATCCC		
	ORF Start: ATG at 36	ORF Stop: TGA at 1068	
	SEQ ID NO: 34	344 aa	MW at 37958.7 Da
NOV6b, CG108782-02 Protein Sequence	MLSLKLPQLLQVHQVPRVFWEDGIMSGYRRPTSSALDCVLSSFQMTNETVNIWTHFLP TWYFLWRLALAGGPGFRAEPYHWPPLLVLFPACLYPFASCCAHTFSSMSPRMRHICY FLDYGALSLYSLGCAFPYAAYSMPASWLHGHLHQFFVPAALNSFLCTGLSCYSRFLE LESPGLSKVLRTGAFAYPFLFDNLPLFYRLGLCWGRGHGCGQEALSTSHGYHLFCALL TGFLFASHLPERLAPGRFDYIGHSHQLFHHICAVLGTGHFQLEAVLADMGSRRRAWLATQE PALGLAGTVATLVLAAGNLLIIAAFTATLLRAPGTCPLLQGGPLEGGTQAKQQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b.		
Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..344	311/344 (90%)
	1..344	311/344 (90%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

<b>Table 6D. Geneseq Results for NOV6a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV6a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB11063	Human secreted protein homologue, SEQ ID NO:1433 - Homo sapiens, 96 aa. [WO200157188-A2, 09-AUG-2001]	176..271 1..96	96/96 (100%) 96/96 (100%)	2e-54
ABB89827	Human polypeptide SEQ ID NO 2203 - Homo sapiens, 284 aa. [WO200190304-A2, 29-NOV-2001]	57..243 36..179	102/190 (53%) 105/190 (54%)	2e-41
AAG01602	Human secreted protein, SEQ ID NO: 5683 - Homo sapiens, 87 aa. [EP1033401-A2, 06-SEP-2000]	1..61 1..61	59/61 (96%) 60/61 (97%)	2e-28
AAG01600	Human secreted protein, SEQ ID NO: 5681 - Homo sapiens, 87 aa. [EP1033401-A2, 06-SEP-2000]	1..61 1..61	59/61 (96%) 60/61 (97%)	2e-28
AAY35973	Extended human secreted protein sequence, SEQ ID NO. 222 - Homo sapiens, 346 aa. [WO9931236-A2, 24-JUN-1999]	14..283 37..301	82/271 (30%) 126/271 (46%)	5e-28

- In a BLAST search of public sequence databases, the NOV6a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 6E.

<b>Table 6E. Public BLASTP Results for NOV6a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV6a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9BGW7	Hypothetical 25.8 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 238 aa.	107..344 1..238	229/238 (96%) 230/238 (96%)	e-135
Q9H621	CDNA: FLJ22672 fis, clone HSI09265 - Homo sapiens (Human), 206 aa.	139..344 1..206	205/206 (99%) 205/206 (99%)	e-119
Q9NXX6	CDNA FLJ20190 fis, clone COLF0714 - Homo sapiens (Human), 330 aa.	1..324 1..321	166/324 (51%) 215/324 (66%)	1e-96
Q9DCU0	0610010115Rik protein - Mus musculus (Mouse), 330 aa.	1..324 1..321	171/324 (52%) 217/324 (66%)	2e-96

Q9DA71	1700019B16Rik protein - Mus musculus (Mouse), 354 aa.	5..322 32..342	104/321 (32%) 151/321 (46%)	7e-34
--------	-------------------------------------------------------	-------------------	--------------------------------	-------

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
UPF0073	33..276	70/292 (24%) 152/292 (52%)	1.5e-09

### Example 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 35	1441 bp	
NOV7a, CG108801-01 DNA Sequence	<div>GGCAGCCGCTTCGGCGCCCCGGCCCCGCGGCCAGCTAGGGGCGGCCCCGCGCTCCCTCA</div> <div>CGGCCCTTCGGCGGCGCCCGTTCGGATCCGGCCTCTCTCTGCGCCCCGGGGCGCGCCAC</div> <div>CTCCCCGCCGGAGGTGTCCACGCGTCCGGCCGTCCATCCGTCCGTCCCTCCTGGGGCC</div> <div>GGCGCTGACCATGCCAGCGGCTGCCGCTGCCTGCATCTCGTGTGCCTGTTGTGCATT</div> <div>CTGGGGGCTCCCGGTGAGCCTGTCCGAGCCGATGACTGCAGCTCCCACTGTGACCTGG</div> <div>CCCACGGCTGCTGTGCACCTGACGGCTCCTGCAGGTGTGACCCGGGCTGGGAGGGGCT</div> <div>GCACTGTGAGCGCTGTGTGAGGATGCCTGGCTGCCAGCACGGTACCTGCCACCAGCCA</div> <div>TGGCAGTGCATCTGCCACAGTGGCTGGGCAGGCAAGTTCTGTGACAAAGATGAACATA</div> <div>TCTGTACCACGCAAGTCCCCTGCCAGAATGGAGGCCAGTGCATGTATACGGGGCGG</div> <div>TGAGTACCATTGTGTGTGCTTACCAGGCTTCCATGGGCGTGACTGCGAGCGCAAGGCT</div> <div>GGACCCGTGTGAACAGGCAGGCTCCCCATGCCGCAATGGCGGGCAGTGCCAGGACGACC</div> <div>AGGGCTTTGCTCTCAACTTCACGTGCCGCTGCTTGGTGGGCTCTGTGGGTGCCGCTG</div> <div>TGAGGTAAATGTGGATGACTGCCTGATGCGGCCCTTGTGCTAACGGTGCCACCTGCCTT</div> <div>GACGGCATAAACCGCTTCTCCTGCCTCTGTCTTGAGGGCTTTGCTGGACGCTTCTGCA</div> <div>CCATCAACCTGGATGACTGTGCCAGCCGCCCATGCCAGAGAGGGGGCCGCTGTGCGGA</div> <div>CCGTGTCCACGACTTCGACTGCCTCTGCCCCAGTGGCTATGGTGGCAAGACCTGTGAG</div> <div>CTTGTCTTACCTGTCCAGACCCCCAACCAAGTGGACAGCCCTCTAGGGCCACCT</div> <div>CAGCTGTAGTGGTACCTGCCACGGGGCCAGCCCCCACAGCGCAGGGGCTGGTCTGCT</div> <div>GCGGATCTCAGTGAAGGAGGTGGTGCAGGCAAGAGGCTGGGCTAGGTGAGCCTAGC</div> <div>TTGGTGGCCCTGGTGGTGTGTTGGGGCCCTCACTGCTGCCCTGGTTCTGGCTACTGTGT</div> <div>TGCTGACCCTGAGGGCCTGGCGCCGGGGTGTCTGCCCTCCTGGACCCTGTTGCTACCC</div> <div>TGCCCCACACTATGCTCCAGCGTGCCAGGACCAGGAGTGTCAGGTTAGCATGCTGCCA</div> <div>GCAGGGCTCCCCCTGCCACGTGACTTGCCCCCTGAGCCTGGAAAGACCACAGCACTGT</div> <div>GATGGAGGTGGGGGCTTTCTGGCCCCCTTCTCACCTCTTCCACCCCTCAGACTGGAG</div> <div>TGGTCCGTTCTACCACCCTTCAGCTTGGGTACACACACAGAAGGGCGA</div>		
	ORF Start: ATG at 185	ORF Stop: TGA at 1334	
	SEQ ID NO: 36	383 aa	MW at 40487.0 Da
NOV7a, CG108801-01 Protein Sequence	<div>MPSGCRCLHLVCLLCILGAPGQPVRRDDCSSHCDLAHGCCAPDGSCRPDGPWEGLHCE</div> <div>RCVRMPGCQHGTCHQPWQCICHSGWAGKFCDEHICTTQSPCQNGGQCMYDGGGEYH</div> <div>CVCLPGFHFGRDCERKAGPCEQAGSPCRNGGQCQDDQGFALNFTCRCLVGSVGARCEVN</div> <div>VDDCLMRPCANGATCLDGINRFSCLCEPFGAGRFCTINLDDCASRPCQRGARCRDRVH</div> <div>DFDCLCPSGYGGKTCELVLPVPDPPTTVDTPLGPTSAVVVPATGPAPHSAGAGLLRIS</div>		

	VKEVVRQEAGLGEPSSLVALVVFALTAALVLATVLLTLRAWRRRGVCPGPCCYPAPHYAPACQDQECQVSMLPAGLPLPRDLPPPEPGKTTAL		
	SEQ ID NO: 37	1348 bp	
NOV7b, CG108801-02 DNA Sequence	GGCAGCCGCTTCGGCGCCCCGGCCCCGGGCCAGCTAGGGGCGGCCCCGCGCTCCCTCACGCCCCCTCGGCGGCGCCCCGTTCGGATCCGGCCTCTCTCTGCGCCCCGGGGCGCGCCACTCCCCGCGCGGAGGTGTCCACGCGTCCGGCCGTCCATCCGTCCGTCCCTCCTGGGGCGCGCTGACCATGCCAGCGGCTGCCGCTGCCTGCATCTCGTGTGCCTGTTGTGCATTCTGGGGGCTCCCGGTACGCCTGTCCGAGCCGATGACTGCAGCTCCCACTGTGACCTGGCCCACGGGTGCTGTGCACCTGACGGCTCCTGCAGGTGTGACCCGGCTGGGAGGGGCTGCACGTGAGCGCTGTGTGAGGATGCCTGGCTGCCAGCACGGTACCTGCCACCAGCCATGGCAGTGCATCTGCCACAGTGGCTGGGCAGGCAAGTTCTGTGACAAAGGCTTCCATGGCGTGACTGCGAGCGCAAGGCTGGACCCTGTGAACAGGCAGGCTCCCCATGCCGCAATGGCGGGCAGTGCCAGGACGACCAGGGCTTTGCTCTCAACTTCACGTGCCGCTGCTTG GTGGGCTCTGTGGGTGCCCGCTGTGAGGTAAATGTGGATGACTGCCTGATGCGGCCTTGTGCTAACGGTGCCACCTGCCTTGACGGCATAAACCGCTTCTCCTGCCTCTGTCTGAGGGCTTTGCTGGACGCTTCTGCACCATCAACCTGGATGACTGTGCCAGCCGCCCATGC CAGAGAGGGGCCCGCTGTCCGGACCGTGTCCACGACTTCGACTGCCTTCGCCCCAGTG GCTATGGTGGCAAGACCTGTGAGCTTGCTTACCTGTCCAGACCCCCCAACCACAGT GGACACCCCTCTAGGGCCCACCTCAGCTGTAGTGGTACCTGCCACGGGGCCAGCCCC CACAGCGCAGGGGCTGGTCTGCTGCGGATCTCAGTGAAGGAGGTGGTGGGAGGCAAG AGGCTGGGCTAGGTGAGCCTAGCTTGGTGGCCCTGGTGGTGTGTTGGGGCCCTCACTGC TGCCCTGGTTCTGGCTACTGTGTTGCTGACCCTGAGGGCCTGGCGCCGGGGTGTCTGC CCTCCTGGACCCTGTTGCTACCCTGCCCCACACTATGCTCCAGCGTGCCAGGACCAGG AGTGTCAAGTTAGCATGCTGCCAGCAGGGCTCCCCCTGCCACGTGACTTGCCCCCTGA GCCTGGAAAGACCACAGCACTGTGATGGAGGTGGGGGCTTTCTGGCCCCCTTCCTCAC CTCTCCACCCCTCAGACTGGAGTGGTCCGTTCTCACCACCCTTCAGCTTGGGTACAC ACACAGAAGGGCGA		
	ORF Start: ATG at 185	ORF Stop: TGA at 1241	
	SEQ ID NO: 38	352 aa	MW at 37158.3 Da
NOV7b, CG108801-02 Protein Sequence	MPSGCRCLHLVCLLCILGAPGQPVRRADDCSSHCDLAHGCCAPDGSCRCDPGWEGLHCE RCVRMPGCQHGTCHQPWQCICHSGWAGKFCDKGFHGRDCERKAGPCEQAGSPCRNGGQ CQDDQGFALNFTCRCLVGSVGARCEVNVDCLMRPCANGATCLDGINRFSCLCPEGFA GRFCTINLDDCASRPCQRGARCRDRVHDFDCLCPSGYGGKTCELVLVPDPPTTVDTP LGPTSAVVVPATGPAPHYSAGAGLLRISVKEVVRQEAGLGEPSSLVALVVFALTAALV LATVLLTLRAWRRRGVCPGPCCYPAPHYAPACQDQECQVSMLPAGLPLPRDLPPPEPGKTTAL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	1..383 1..352	296/383 (77%) 296/383 (77%)

Further analysis of the NOV7a protein yielded the following properties shown in Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic

	reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 27 and 28

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7D.

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG67516	Amino acid sequence of a human secreted polypeptide - Homo sapiens, 383 aa. [WO200166690-A2, 13-SEP-2001]	1..383 1..383	382/383 (99%) 382/383 (99%)	0.0
AAE01167	Human gene 4 encoded secreted protein HKAAV61, SEQ ID NO:68 - Homo sapiens, 383 aa. [WO200134768-A2, 17-MAY-2001]	1..383 1..383	382/383 (99%) 382/383 (99%)	0.0
AAE13632	Human preadipocyte factor-1-like protein - Homo sapiens, 383 aa. [WO200157233-A2, 09-AUG-2001]	1..383 1..383	381/383 (99%) 381/383 (99%)	0.0
AAE13639	Human preadipocyte factor-1-like protein fragment #1 - Homo sapiens, 357 aa. [WO200157233-A2, 09-AUG-2001]	27..383 1..357	356/357 (99%) 356/357 (99%)	0.0
AAE13641	Wheat germ agglutinin #1 found in human Pref-1-like protein - Triticum aestivum, 177 aa. [WO200157233-A2, 09-AUG-2001]	57..233 1..177	176/177 (99%) 176/177 (99%)	e-116

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7E.

Table 7E. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BQ54	Hypothetical 21.3 kDa protein (Unknown) (Protein for MGC:2487) - Homo sapiens (Human), 204 aa.	180..383 1..204	204/204 (100%) 204/204 (100%)	e-122
O70534	ZOG protein - Rattus norvegicus (Rat), 383 aa.	10..327 7..324	127/325 (39%) 171/325 (52%)	2e-67

Q62779	Preadipocyte factor 1 - Rattus norvegicus (Rat), 383 aa.	10..325 7..322	127/323 (39%) 169/323 (52%)	9e-67
Q925U3	Dlk (Delta like) (Delta-like) - Mus musculus (Mouse), 385 aa.	10..327 7..326	126/327 (38%) 170/327 (51%)	1e-64
Q09163	Delta-like protein precursor (DLK) (Preadipocyte factor 1) (Pref-1) (Adipocyte differentiation inhibitor protein) [Contains: Fetal antigen 1 (FA1)] - Mus musculus (Mouse), 385 aa.	10..327 7..326	126/327 (38%) 170/327 (51%)	1e-64

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF	60..88	11/47 (23%) 23/47 (49%)	0.0016
EGF	95..128	16/47 (34%) 30/47 (64%)	8e-08
EGF	135..171	15/47 (32%) 23/47 (49%)	0.0003
EGF	178..209	13/47 (28%) 26/47 (55%)	6.1e-09
EGF	216..247	14/47 (30%) 24/47 (51%)	5.2e-06

### Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis		
	SEQ ID NO: 39	2484 bp
NOV8a, CG109717-01 DNA Sequence	GGATCTCAGCACTCTGACCCAAGGGGAAGCATGTCTGAAGAAAGGCCGAGCAAGGGCG AGAAGCCCGAGATGGAGACGGACGCGGTGCAGATGGCCAACGAGGAGCTGCGGGCCAA GCTGACCAGCATTAGATCGAGTTCCAGCAGGAAAAAAGCAAGGTGGGCAAAGTGCAGC GAGCGGCTGCAGGAGGCGAAGCTGGAGCGCGAGCAGGAGCAGCGACGGCACACGGCCT ACATTTTCGGAGCTCAAGGCCAAGCTGCATGAGGAGAAGACCAAGGAGCTGCAGGCGCT GCGCGAGGGGCTCATCCGGCAGCACGAGCAGGAGGCGGCGCGCACCGCCAAGATCAAG GAGGGCGAGCTGCAGCGGCTGCAGGCCACGCTGAACGTGCTGCGCGACGGCGCGGCCG ACAAGGTCAAGACGGCGCTGCTGACCGAGGCGCGGAGGAGGCGCGCAGGGCCTTCGA TGGAGAGCGCCTGCGGCTGCAGCAGGAGATCCTGGAGCTCAAGGCAGCGCGCAAGCAG GCAGAGGAGGCGCTCAGTAAGTGCATGCAGGCTGACAAGACCAAGGCAGCCGACCTGC GTGCCGCTACAGGCGCACCAAGACGAGGTGCACCGCATCAAGCGCAGTGCAGGCG CGACATCCGAGGCTGATGGATGAGATCAAAGGGAAGACCGTGTGATTCTGGCCTTG	

	GAGAAGGAACCTTGGCGTGCAGGCTGGGCAGACCCAGAAGCTGCTTCTGCAGAAAGAGG CTTTGGATGAGCAGCTGGTTTCAGGTCAAGGAGGCCGAGCGGCACCACAGTAGTCCAAA GAGAGAGCTCCCGCCCGGGATCGGGGACATGGTGGAGCTCATGGGCGTCCAGGATCAA CATATGGACGAGCGAGATGTGAGGCGATTTCAACTAAAAATTGCTGAACTGAATTTCAG TGATACGGAAGCTGGAAGACAGAAATACGCTGTTGGCAGATGAGAGGAATGAACTGCT GAAACGCTCACGAGAGACCGAGGTTTCAGCTGAAGCCCCCTGGTGGAGAAGAACAAGCGG ATGAACAAGAAGAATGAGGATCTGTTGCAGAGTATCCAGAGGATGGAGGAGAAAATCA AGAACCTCACGCGGGAAAACGTGGAAATGCTGTCAGCGCAGGCGTCTCTGAAGCGGCA TACCTCCTTGAATGACCTCAGCCTGACGAGGGATGAGCAGGAGATCGAGTCTCTGAGG CTGCAGGTGCTGGAGCAGCAGCACGTCATTGACGACCTCTCAC'TGGAGAGAGAACGGC TGTTGCGCTCCAAAAGGCATCGAGGGAAAAGTCTGAAACCGCCCCAAGAAGCATGTTGT GGAGACATTTTTTGGATTTGATGAGGAGTCTGTGGACTCAGAAACGTTGTCCGAAACA TCCTACAACACAGACAGGACAGACAGGACCCAGCCACGCCCCGAAGAAGACTTGGACG ATAAGGCCACAGCCCCGAGAGGAGGCTGACCTGCGCTTCTGCCAGCTGACCCGGGAGTA CCAGGCCCTGCAACGCGCCTACGCCCTGCTCCAGGAGCAGGTGGGAGGCACGCTGGAC GCTGAGAGGGAGGCCCGGACTCGGGAGCAGCTACAAGCTGATCTGCTGAGGTGTGAGG CCAAAATCGAAGATTTGGAGAAGTTACTGGTTGAGAAGGGACAGGTGAGCAGGAGTGA TATGGAAGAGAACCAGCTGAAGAATGAAATGCAAGACGCCAAGGATCAGAACGAGCTG TTAGAATTCAGAGTGCTAGAACTCGAAGAGAGAGAGAGGAGGTGCGCAGCATTTAACC TCCAAATCACACCTTCCCCGAGAACCACAGCAGCGCTCTCCAGCTGTTCTGTACCA GGAAGGAGTTAAGGATGTGAATGTTTCTGAACTTATGAAGAAATTAGATATCCTTGGC GATAACGGGAATTTGAGAAATGAAGAACAGGTTGCAATAATCCAAGCTGGAAGTGTGC TTGCCCTGTGTGAAAAGTGGCTGAAGCAAATAGAGGGGACCGAGGCCGCCCTGACCCA GAAGATGCTGGACCTGGAGAAGGAGCAGGACCTGTTTCAGCAGGCAGAAGGGCTACCTG GAAGAGGAGCTCGACTACCGGAAGCAAGCCCTTGACCAGGCTTACCTGAAAATCCAAG ACCTGGAGGCCACACTGTACACAGCGCTGCAGCAGGAGCCGGGCGGAGGGCCGTGA GGCGCTGAGCGAGGGCCAGCGGGAGGACCTGCAGGCTGCTGTGAAAAGGTGCGCAGG CAGATCCTCAGGCAGAGCCGCGAGTTTCGACAGCCAGATCCTGCGGGAGCGCATGGAGC TGCTGCAGCAGGCCCCAGCAGAGAATCCGAGAACTGGAGGACAAACTGGAGTTTCAGAA GCGGCACCTGAAAGAACTGGAGGAAAAGTTTTTGTTCCTTTTTTTGTTTTTCTACTA GCATTCATTCTGTGGCCTTGATGACTTCAGTGAGCCAAGAAGCTCGGGT		
	ORF Start: ATG at 31	ORF Stop: TGA at 2455	
	SEQ ID NO: 40	808 aa	MW at 94479.1 Da
NOV8a, CG109717-01 Protein Sequence	MSKKGRSKGEKPEMETDAVQMANEELRAKLTSIQIEFQQEKSQVGLRERLQEAALER EQEQRHTAYISELKAKLHEEKTKEQLALREGLIRQHEQEAARTAKIKEGELQRLQAT LNVLRDGAADKVKTAALLTEAREEARAFDGERLRLQQEILELKAARKQAEELSNMCQ ADKTKAADLRAAYQAHQDEVHRIKRECERDIRRLMDEIKGKDRVILALEKELGVQAGQ TQKLLQKEALDEQLVQVKEAERHHSSPKRELPPGIGDMVELMGVQDQHMDERDVRRF QLKIAELNSVIRKLEDNRNTLLADERNELLKRSRETEVQLKPLVEKNRNMKNKEDLLQ SIQRMEEKIKNLTRENVEMLSAQASLKRHTSLNDLSLTRDEQIEFLRLQVLEQQHVI DDLSELERERLLRSKRHRGKSLKPPKKHVETFFGFDEESVDSETLSETSYNTDRDRT PATPEEDLDDKATAREEADLRFQLTREYQALQRAYALLQEQQVGTLDAREARTREQ LQADLLRCQAKIEDLEKLLVEKGQVSRSDMEENQLKNEMQDAKDQNELLEFRVLELEE RERRSPAFNLQITTFPENHSSALQLFCHQEGVDVNVSELMKKLDILGDNGNLRNEEQ VAIIQAGTVLALCEKWLKQIEGTEAALTQKMLDLEKEQDLFSRQKGYLEELDYRKQA LDQAYLKIQDLEATLYTALQQEPGRRAGEALSEGQREDLQAAVEKVRRIILRQSREFD SQILRERMELLOQAQQRIRELEDKLEFQKRHLKELEEKFLFLFFSLAFILWP		

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial inner membrane
SignalP	No Known Signal Sequence Predicted



analysis:	
-----------	--

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8C.

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB04608	Human xylose isomerase 43 protein SEQ ID NO:2 - Homo sapiens, 387 aa. [CN1307130-A, 08-AUG-2001]	173..582 1..366	323/431 (74%) 332/431 (76%)	e-162
AAB42436	Human ORFX ORF2200 polypeptide sequence SEQ ID NO:4400 - Homo sapiens, 241 aa. [WO200058473-A2, 05-OCT-2000]	194..431 1..241	238/241 (98%) 238/241 (98%)	e-128
AAM85650	Human immune/haematopoietic antigen SEQ ID NO:13243 - Homo sapiens, 388 aa. [WO200157182-A2, 09-AUG-2001]	445..808 4..388	238/390 (61%) 298/390 (76%)	e-124
ABB61173	Drosophila melanogaster polypeptide SEQ ID NO 10311 - Drosophila melanogaster, 1690 aa. [WO200171042-A2, 27-SEP-2001]	6..788 423..1263	179/877 (20%) 360/877 (40%)	4e-24
AAY30795	A human trichohyalin (TRHY) protein - Homo sapiens, 1898 aa. [US958752-A, 28-SEP-1999]	24..794 258..991	171/792 (21%) 345/792 (42%)	5e-24

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96N16	CDNA FLJ31564 fis, clone NT2RI2001450, weakly similar to trichohyalin - Homo sapiens (Human), 626 aa.	1..582 1..605	575/606 (94%) 578/606 (94%)	0.0
T00331	hypothetical protein KIAA0555 - human, 799 aa.	1..808 1..799	530/812 (65%) 656/812 (80%)	0.0
Q96AA8	Hypothetical protein KIAA0555 - Homo sapiens (Human), 810 aa.	1..792 1..804	513/817 (62%) 641/817 (77%)	0.0

Q9CU41	6330417G02Rik protein - Mus musculus (Mouse), 437 aa (fragment).	1..418 1..435	262/436 (60%) 333/436 (76%)	e-139
Q9BGP2	Hypothetical 23.9 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 201 aa.	609..808 2..201	148/200 (74%) 177/200 (88%)	1e-79

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8E.

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

### Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis		
	SEQ ID NO: 41	3040 bp
NOV9a, CG110477-01 DNA Sequence	ACAATGATGGGGCTCTTCCCCAGAACTACAGGGGCTCTGGCCATCTTCGTGGTAGTCA TATTGGTTCATGGAGAATTGCGAATAGAGACTAAAGGTCAATATGATGAAGAAGAGAT GACTATGCAACAAGCTAAAAGAAGGCCAAAACGTGAATGGGTGAAATTTGCCAAACCC TGCAGAGAAGGAGAAGATAACTCAAAAAGAAACCCAATTGCCAAGATTACTTCAGATT ACCAAGCAACCCAGAAAATCACCTACCGAATCTCTGGAGTGGGAATCGATCAGCCGCC TTTTGAATCTTTGTTGTTGACAAAAACACTGGAGATATTAACATAACAGCTATAGTC GACCGGGAGGAACTCCAAGCTTCCTGATCACATGTCGGGCTCTAAATGCCCAAGGAC TAGATGTAGAGAAACCACTTATACTAACGGTTAAAATTTTGGATATTAATGATAATCC TCCAGTATTTTCACAACAAATTTTCATGGGTGAAATTGAAGAAAATAGTGCCTCAGAC TCACTGGTGTATGATACTAAATGCCACAGATGCAGATGAACCAAACCACTTGAATTCTA AAATTGCCTTCAAAATTGTCTCTCAGGAACCAAGCAGGCACACCCATGTTCCCTCCTAAG CAGAAACACTGGGGAAGTCCGTACTTTGACCAATTCTCTTGACCGAGAGCAAGCTAGC AGCTATCGTCTGGTTGTGAGTGGTGCAGACAAAGATGGAGAAGGACTATCAACTCAAT GTGAATGTAATATTAAAGTGAAGATGTCAACGATAACTTCCCAATGTTTAGAGACTC TCAGTATTCAGCACGTATTGAAGAAAATATTTTAAGTTCTGAATTACTTCGATTTCAA GTAACAGATTTGGATGAAGAGTACACAGATAATTGGCTTGCAGTATATTTCTTTACCT CTGGGAATGAAGGAAATTGGTTTGAATAACAACTGATCCTAGAACTAATGAAGGCAT CCTGAAAGTGGTGAAGGCTCTAGATTATGAACAACCTACAAAGCGTGAAACTTAGTATT GCTGTCAAAAACAAAGCTGAATTTACCAATCAGTTATCTCTCGATACCGAGTTCAGT CAACCCAGTCACAATTCAGGTAATAAATGTAAGAGAAGGAATTGCATTCCGTCCTGCT TTCCAAGACATTTACTGTGCAAAAAGGCATAAAGTAGCAAAAATTTGGTGGATTATATC CTGGGAACATATCAAGCCATCGATGAGGACACTAACAAAGCTGCCTCAAATGTCAAGT ATGTCATGGGACGTAACGATGGTGGATACCTAATGATTGATTCAAAAACCTGCTGAAAT CAAATTTGTCAAAAATATGAACCGAGATTCTACTTTCATAGTTAACAAAACAATCACA GCTGAGGTTCTGGCCATAGATGAATACACGGGTAAACTTCTACAGGCACGGTATATG TTAGAGTACCCGATTTCAATGACAATTGTCCAACAGCTGTCTCGAAAAAGATGCAGT TTGCAGTTCTTCACCTTCCGTGGTTGTCTCCGCTAGAACACTGAATAATAGATACACT GGCCCCATACATTTGCACTGGAAGATCAACCTGTAAAGTTGCCTGCCGTATGGAGTA TCACAACCCCTCAATGCTACCTCGGCCCTCCTCAGAGCCCAGGAACAGATACCTCTGG AGTATACCACATCTCCCTGGTACTTACAGACAGTCAGAACAATCGGTGTGAGATGCCA CGCAGCTTGACACTGGAAGTCTGTCAAGTGTGACAACAGGGGCATCTGTGGAACCTCTT	

	ACCCAACCACAAGCCCTGGGACCAGGTATGGCAGGCCGCACTCAGGGAGGCTGGGGCC TGCCGCCATCGGCCTGCTGCTCCTTGGTCTCCTGCTGCTGCTGGTGGCCCCCCTTCTG CTGTTGACCTGTGACTGTGGGGCAGGTTCTACTGGGGAGTGACAGGTGGTTTTATCC CAGTTCCTGATGGCTCAGAAGGAACAATTCATCAGTGGGGAATTGAAGGAGCCCATCC TGAAGACAAGGAAATCACAAATATTTGTGTGCCTCCTGTAACAGCCAATGGAGCCGAT TTCATGGAAAGTTCTGAAGTTTGTACAAATACGTATGCCAGAGGCACAGCGGTGGAAG GCACTTCAGGAATGGAATGACCACTAAGCTTGGAGCAGCCACTGAATCTGGAGGTGC TGCAGGCTTTGCAACAGGGACAGTGTGAGGAGCTGCTTCAGGATTCGGAGCAGCCACT GGAGTTGGCATCTGTTCCCTCAGGGCAGTGTGGAACCATGAGAACAAGGCATTCACATG GAGGAACCAATAAGGACTACGCTGATGGGGCGATAAGCATGAATTTTCTGGACTCCCTA CTTTTCTCAGAAAGCATTTCCTGTGCGGAGGAAGACGATGGCCAGGAAGCAAATGAC TGCTTGTTGATCTATGATAATGAAGGCGCAGATGCCACTGGTTCTCCTGTGGGCTCCG TGGGTTGTTGCAGTTTTATTGCTGATGACCTGGATGACAGCTTCTTGGACTCACTTGG ACCCAAATTTAAAAAAGTTCAGAGATAAGCCTTGGTGTGATGGTGAAGCAAAGAA GTTTCAGCCACCCTCTAAAGACAGCGGTTATGGGATTGAATCCTGTGGCCATCCCATAG AAGTCCAGCAGACAGGATTTGTAAAGTGCCAGACTTTGTGAGGAAGTCAAGGAGCTTC TGCTTTGTCCACCTCTGGGTCTGTCCAGCCAGCTGTTTCCATCCCTGACCCCTCTGCAG CATGGTAACTATTTAGTAACGGAGACTTACTCGGCTTCTGGTTCCTCGTGCAACCTT CCACTGCAGGCTTTGATCCACTTCTCACACAAAATGTGATAGTGACAGAAAGGTGAT CTGTCCCATTTCAGTGTTCCTGGCAACCTAGCTGGCCCAACGCAGCTACGAGGGTCA CATACTATGCTCTGTACAGAGGATCCTTGCTCCCGTCTAATATGACCAGAATGAGCTG GAATACCACACTGACCAAATCTGG		
	ORF Start: ATG at 4	ORF Stop: TGA at 3001	
	SEQ ID NO: 42	999 aa	MW at 107518.8 Da
NOV9a, CG110477-01 Protein Sequence	MMGLFPRRTGALAI FVVVILVHGE LRIETKGQYDEEEMTMQQAKRRQKREWVKFAKPC REGEDNSKRNP IAKITS DYQATQKITYRISGVGIDQPPFGIFVVDKNTGDINITAIVD REETPSFLITCRALNAQGLDVEKPLILTVKILDINDNPPVFSQQIFMGEIEENSASDS LVMILNATDADEPNHLNSKIAFKIVSQEPAGTPMFLLSRNTGEVRTLTNSLDREQASS YRLVVSGADKDGEGLS TQCECN IKVKDVNDNFPMFRDSQYSARIEENILSSELLRFQV TDLDEEYTDNLAVYFFTSGNEGNWF EIQTDPR TNEGILKVVKALDYEQLQSVKLSIA VKNKAEFHQSVISRYRVQSTPVTIQVINVREGIAFRPASKTFTVQKGISSKKLV DYIL GTYQAIDEDTNKAASN VKYVMGRNDGGYLMIDSKTAEIKFVKNMNRDSTFIVNKTITA EVLAIDEYTGKTSTGT VYVRVPDFDNCPTAVLEKDAVCSSSPVVVSARTLNNRYTG PYTFALEDQPVKLP AVWSITTLNATSALLRAQEQIPPGVYHISLVLTDSQNNRCEMPR SLTLEVCQCDNRGICGTSYPTTSPGTRYGRPHSGRLGPAAIGLLLLLGLLLL LVAPLLL LTCDCGAGSTGGVTGGFIPVPD GSEGTIHQWGIEGAHPEDKEITNICVPPVTANGADF MESSEVCTNTYARGTAVEGTSGMEMTTKLGAATESGGAAGFATGT VSGAASGFGAATG VGICSSGQSGTMRTRHSTGGTNKDYADGAISMNFLDSYFSQKAFACAEEDDGQEANDC LLIYDNEGADATGSPVGSVGC CSFIADDLDDSF LDSLGP KFKKLAEISLGVDGEGKEV QPPSKDSGYGIESCGHP IEVQQTGFVKCQTLSGSQGASALSTSGSVQPAVSI PDPLQH GNYLVTETYSAGSLVQ PSTAGFDPLLTQNVIVTERVICPISSVPGNLAGPTQLRGSH TMLCTEDPCSRLI		

Further analysis of the NOV9a protein yielded the following properties shown in Table 9B.

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9C.

<b>Table 9C. Geneseq Results for NOV9a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV9a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAU78054	Human desmoglein 3 (pemphigus vulgaris antigen) protein sequence - Homo sapiens, 999 aa. [WO200210767-A2, 07-FEB-2002]	1..999 1..999	996/999 (99%) 998/999 (99%)	0.0
ABG12435	Novel human diagnostic protein #12426 - Homo sapiens, 1014 aa. [WO200175067-A2, 11-OCT-2001]	1..999 16..1014	996/999 (99%) 998/999 (99%)	0.0
ABG12435	Novel human diagnostic protein #12426 - Homo sapiens, 1014 aa. [WO200175067-A2, 11-OCT-2001]	1..999 16..1014	996/999 (99%) 998/999 (99%)	0.0
AAR30742	Human pemphigus vulgaris 130kD antigen - Homo sapiens, 999 aa. [USN7798918-N, 15-DEC-1992]	1..999 1..999	996/999 (99%) 998/999 (99%)	0.0
AAW07908	Pemphigus vulgaris antigen protein extracellular region - Homo sapiens, 614 aa. [JP08188540-A, 23-JUL-1996]	2..615 1..614	610/614 (99%) 612/614 (99%)	0.0

In a BLAST search of public sequence databases, the NOV9a protein was found to

- 5 have homology to the proteins shown in the BLASTP data in Table 9D.

<b>Table 9D. Public BLASTP Results for NOV9a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV9a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P32926	Desmoglein 3 precursor (130 kDa pemphigus vulgaris antigen) (PVA) - Homo sapiens (Human), 999 aa.	1..999 1..999	996/999 (99%) 998/999 (99%)	0.0
O35902	Desmoglein 3 precursor (130 kDa pemphigus vulgaris antigen homolog) - Mus musculus (Mouse), 993 aa (fragment).	1..998 1..993	729/1018 (71%) 832/1018 (81%)	0.0
Q02413	Desmoglein 1 precursor (Desmosomal glycoprotein 1) (DG1) (DGI) (Pemphigus foliaceus antigen) - Homo sapiens (Human), 1049 aa.	5..992 5..896	429/1003 (42%) 581/1003 (57%)	0.0

Q8R517	Desmoglein 2 - Mus musculus (Mouse), 1122 aa.	46..972 51..977	393/960 (40%) 559/960 (57%)	0.0
Q14126	Desmoglein 2 precursor (HDGC) - Homo sapiens (Human), 1117 aa.	46..972 45..973	376/963 (39%) 558/963 (57%)	e-177

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

Table 9E. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cadherin	54..148	23/111 (21%) 68/111 (61%)	6.5e-06
cadherin	162..258	30/110 (27%) 75/110 (68%)	4e-21
cadherin	272..375	33/107 (31%) 88/107 (82%)	1.6e-30
cadherin	388..486	24/113 (21%) 68/113 (60%)	0.00099

#### Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis			
	SEQ ID NO: 43	898 bp	
NOV10a, CG110540-01 DNA Sequence	TAAGATGAATAAAAAACAACAAACCTTCCAGTTTCATAGCCATAAGAAATGCTGCTTTCTCTGAAGTCGGCATTGGGATCTCTGCCAATGCCATGCTCCTTCTCTCCACATCCTCAGTGCTTCTCAAGCACAGGACCAAGCCCGCTGACCTGATCGTTTGTCATGTGGCTCTAATCCATATCATATTGCTGCTACCCACAGAGTTCATAGCTACAGATATTTTTGGGTCTCAGGATTCAGAGGATGACATCAAACATAAGTCAGTTATCTACAGGTACAGGTTGATGAGAGGCCTCTCCATTTCCACCACCTGCCTGCTGAGTATCCTCCCGGCCATCACCTGCAGCCCCAGAAGCTCCTGTTTGGCAGTGTTCAAAGATTCTCACATCACCAACCACGTTGCTTTCTCTTCCTATGGGTCTTCCACATATCCATTAGTGACAGCTTCTTAGTCTCCACTCTTCCCATCAAAAATCTGGCCTCAAATAGCCTTACATTTGTCACTCAATCCTGCTCTGCTGGGATCCTGAGTTGCTTCCTTGAGCAGACAATTTTCACTGATGACATTTTCAGGATGTCTCCCTTGCAGGGCTCACGGCCCCCTCCAGTGGATACATGGTGATTCTCTTGTTCCAGGCGTAACAGGCAGTCCCAGCATTTTTCACAGCACCAACCTTTCTCCAAAAGCACCCCCAGAAAAAATGGCCACGCAGACCATTCTTCTGCTCGTGAGTTGCTTTGTGATTGTGTATGTTTTGGACTGTGTTGTGCGCTCCTGCTCAGGACTGGTGTGGAACAGTGATCCAGTCCGTCATCGAGTCCAGATGCTGGTGGACAATGGCTATGCCACCATCAGTCCTTCAGTGCTAGTCAGTACTGAAAAATGAATGATCAAA		
	ORF Start: ATG at 5	ORF Stop: TGA at 887	
	SEQ ID NO: 44	294 aa	MW at 32551.7 Da
NOV10a, CG110540-01	MNKNNKPSSFIAIRNAAFSEVGIGISANAMLLLFHILTCLLKHRTKPADLIVCHVALIHIILLPTEFIATDIFGSQDSEDDIKHKSVIYRYRLMRGLSISTTCLLSILPAITCSP		

Protein Sequence	RSSCLAVFKRFSHHQPRCFLFLWVFHISISDSFLVSTLPIKNLASNSLTFVTQSCSAG ILSCFLEQTIIFTLMTFQDVSLAGLTAPSSGYMVILLSRNRNRSQHFSHTNLSPKAPPE KMATQTILLVSCFVIVYVLD CVVASCSGLVWNSDPVRHRVQMLVDNGYATISPSVLV STEK		
	SEQ ID NO: 45	1420 bp	
NOV10b, CG110578-02 DNA Sequence	TGTGGGTCGCTGCTTCCTGGCCCTTCTCCGACCCCGCTCTAGCAGCAGACCTCCTGGG GTCTGTGGGTTGATCTGTGGCCCCTGTGCCTCCGTGTCTCTTTTCGTCTCCCTTCCTCC CGACTCCGCTCCCGGACCAGCGGCCTGACCCTGGGGAAGGATGGTTCCCGAGGTGAG GGTCCTCTCCTCCTTGCTGGGACTCGCGCTGCTCTGGTTCCCCCTGGACTCCCACGCT CGAGCCCGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGA GCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTC AGAGGGCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCTGTCCACTGCCCCCAG CCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAACCTCACACTCCCTCTG GACTCCGGGCCCCACCAAAGTCTGCCAGCACAAACGGGACCATGTACCAACACGGAGA GATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCTCTGC AGCTGCACAGAGGGCCAGATCTACTGCGGCCTCACAACTGCCCCGAACCAGGCTGCC CAGCACCCCTCCCGCTGCCAGACTCCTGCTGCCAGGCCTGCAAAGATGAGGCAAGTGA GCAATCGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTCAGGAT CCATGTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCAGCCCCACTGGCC TCAGCGCCCCCTCTGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCAC AACTGTCAAGATCGTCTTGAAGGAGAAACATAAGAAAGCCTGTGTGCATGGCGGGAAG ACGTACTCCACGGGGAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGCCCCCTTGCCCT GCATCCTATGCACCTGTGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCAC CGAGTACCCCTGCCGTACCCCGAGAAAGTGGCTGGGAAGTGTGCAAGATTTGCCCA GAGGACAAAGCAGACCCTGGCCACAGTGAGATCAGTTCTACCAGGTGTCCCAAGGCAC CGGGCCGGGTCTCTCGTCCACACATCGGTATCCCCAAGCCCAGACAACCTGCGTCGCTT TGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGATCTACCTCTGGAAGCTGGTAAAA GGAATCTTCCACTTGACTCAGATCAAGAAAGTCAGGAAGCAAGACTTCCAGAAACACA TACGCCTCTTCCCTCTTCTGCCCTCCTCCATGCAGGTCACTGGAACGTCTTCTAGCC CAGATCCTGGAGCTGAAGGTCACGGCCA		
	ORF Start: ATG at 158	ORF Stop: TAG at 1388	
	SEQ ID NO: 46	410 aa	MW at 45294.6 Da
NOV10b, CG110578-02 Protein Sequence	MVPEVRVLSSLLGLALLWFPLDSHARARPD MFCLFHGKRYSPGESWHPYLEPQGLMYC LRCTCSEGAHVSCYRLHCPVHCPQPVTEPQQCCPKCPEPHTPSGLRAPPKSCQHNGT MYQHGEIFSAHELFP SRLPNQCVLCSCTEGQIYCGLTTCPEPGCPAPLPLPDSCCQAC KDEASEQSDEEDSVQSLHGVRHPQDPCSSDAGRKRGPPTPAPTGLSAPLSFIPRHRFP KGAGSTTVKIVLKEKHKKACVHGGKTYSHGEVWHPAFRAFGPLPCILCTCEDGRQDCQ RVTCPT EYPCRHPKVKAGCKKICPEDKADPGHSEISSTRCPKAPGRVLVHTSVSPSP DNLRRFALEHEASDLVEIYLWKLVKGI FHLTQIKKVRKQDFQKHIRLFPLLPSSMQVT GTSS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b.		
Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	254..260 138..144	4/7 (57%) 6/7 (85%)

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

<b>Table 10C. Protein Sequence Properties NOV10a</b>	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3331 probability located in mitochondrial inner membrane; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 46 and 47

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10D.

<b>Table 10D. Geneseq Results for NOV10a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV10a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE18646	Human G-protein coupled receptor (GCREC-7) - Homo sapiens, 271 aa. [WO200210387-A2, 07-FEB-2002]	1..294 1..271	258/294 (87%) 258/294 (87%)	e-138
AAW19107	Rat pheromone receptor VN6 - Rattus sp, 310 aa. [WO9714790-A1, 24-APR-1997]	18..293 18..293	140/277 (50%) 181/277 (64%)	7e-67
AAM48284	Pheromone receptor protein VN1-18 - Unidentified, 165 aa. [WO200206333-A1, 24-JAN-2002]	1..125 17..141	125/125 (100%) 125/125 (100%)	6e-66
AAW19104	Rat pheromone receptor VN3 - Rattus sp, 311 aa. [WO9714790-A1, 24-APR-1997]	1..294 2..295	135/295 (45%) 185/295 (61%)	8e-62
AAW19103	Rat pheromone receptor VN1 - Rattus sp, 315 aa. [WO9714790-A1, 24-APR-1997]	1..294 2..295	133/295 (45%) 190/295 (64%)	7e-61

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

<b>Table 10E. Public BLASTP Results for NOV10a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV10a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q8WNV6	Putative pheromone receptor gVIR1 - Capra hircus (Goat), 308 aa.	3..294 2..294	172/293 (58%) 205/293 (69%)	6e-84
Q62855	Pheromone receptor VN6 - Rattus norvegicus (Rat), 310 aa.	18..293 18..293	140/277 (50%) 181/277 (64%)	2e-66

Q9EPA4	VN12 (VOMERONASAL receptor VIRAL) - Mus musculus (Mouse), 303 aa.	1..294 1..294	136/295 (46%) 193/295 (65%)	7e-64
Q8VIC6	Vomeroneasal receptor 1 A8 - Mus musculus (Mouse), 329 aa.	1..294 27..320	136/295 (46%) 193/295 (65%)	7e-64
Q9Z195	Pheromone receptor 1 - Mus musculus (Mouse), 305 aa.	1..294 3..296	136/295 (46%) 193/295 (65%)	7e-64

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

#### Example 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 47	1024 bp	
NOV11a, CG110725-01 DNA Sequence	<u>GACTCGTCTCAGGCCAGTTGCAGCCTTCTCAGCCAAACGCCGACCAAGGAAAACCTCAC</u> <u>TACCATGAGAATTGCAGTGATTGCTTTGCCTCCTAGGCATCACCTGTGCCATACCA</u> <u>GTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAGCTTTACAACAAATACCCAG</u> <u>ATGCTGTGGCCACATGGCTAAACCCTGACCCATCTCAGAAGCAGAATCTCCTAGCCCC</u> <u>ACAGAATGCTGTGTCCTCTGAAGAAACCAATGACTTTAAACAAGAGACCCCTTCCAAGT</u> <u>AAGTCCAACGAAAGCCATGACCACATGGATGATATGGATGATGAAGATGATGATGACC</u> <u>ATGTGGACAGCCAGGACTCCATTGACTCGAACGACTCTGATGATGTAGATGACACTGA</u> <u>TGATTCTCACCAGTCTGATGAGTCTCACCATTCTGATGAATCTGATGAACTGGTCACT</u> <u>GATTTTCCACGACCTGCCAGCAACCGAAGTTTTCAC'TCCAGTTGTCCCCACAGTAG</u> <u>ACACATATGATGGCCGAGGTGATAGTGTGGTTTATGGACTGAGGTCAAATCTAAGAA</u> <u>GTTTCGCAGACCTGACATCCAGTACCCTGATGCTACAGACGAGGACATCACCTCACAC</u> <u>ATGGAAGCGAGGAGTTGAATGGTGCATACAAGGCCATCCCGTTGCCAGGACCTGA</u> <u>ACGCGCCTTCTGATTGGGACAGCCGTGGGAAGGACAGTTATGAAACGAGTCAGCTGGA</u> <u>TGACCAGAGTGCTGAAACCCACAGCCACAAGCAGTCCAAAGTCAGCCGTGAATTCCAC</u> <u>AGCCATGAATTTACAGCCATGAAGATATGCTGGTTGTAGACCCCAAAAGTAAGGAAG</u> <u>AAGATAAACACCTGAAATTTCTGATTTCTCATGAATTAGATAGTGCATCTTCTGAGGT</u> <u>CAATTAAAAGGAGAAAAAATAACAATTTCTCACTTTGCATTTAGTCAAAAGAAAAAAT</u> <u>GCTTTATAGCAAAATGAAAGAGAACATGAAATGCTTCT</u>		
	ORF Start: ATG at 63	ORF Stop: TAA at 933	
	SEQ ID NO: 48	290 aa	MW at 32606.4 Da
NOV11a, CG110725-01 Protein Sequence	MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDVAVATWLNPDPSQKQNLAPQ NAVSSEETNDFKQETLPSKSNESHDMDDDEDHVDSDSIDSNDSDDVDVDDTDD SHQSDESHSHDESDELVTDFPTDLPAVEVFTPVVPTVDYDGRGDSVYGLRSKSKKF RRPDIQYPDATDEDITSHMESEELNGAYKAIPVAQDLNAPSDWDSRGKDSYETSQDLD QSAETHSHKQSKVSREFHSHEFHSHEDMLVVDPKSKEEDKHLKFRISHELDSASSEVN		



	SEQ ID NO: 119		834 bp	
NOV11b, 209934449 DNA Sequence	GGATCCATACCAGTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAGCTTTACAACAAATACCCAG ATGCTGTGGCCACATGGCTAAACCCCTGACCCATCTCAGAAGCAGAATCTCCTAGCCCCACAGAATGCTGT GTCCTCTGAAGAAACCAATGACTTTAAACAAGAGACCCCTCCAAGTAAGTCCAACGAAAGCCATGACCAC ATGGATGATATGGATGATGAAGATGATGATGACCATGTGGACAGCCAGGACTCCATTGACTCGAACGACT CTGATGATGTAGATGACACTGATGATTCTCACCAGTCTGATGAGTCTCACCATTCTGATGAATCTGATGA ACTGGTCACTGATTTTCCACGGACCTGCCAGCAACCGAAGTTTTCACTCCAGTTGTCCCCACAGTAGAC ACATATGATGGCCGAGGTGATAGTGTGGTTTATGGACTGAGGTCAAAATCTAAGAAGTTTCGCAGACCTG ACATCCAGTACCCTGATGCTACAGACGAGGACATCACCTCACACATGGAAGCGAGGAGTTGAATGGTGC ATACAAGGCCATCCCCGTTGCCAGGACCTGAACGCGCCTTCTGATTGGGACAGCCGTGGGAAGGACAGT TATGAAACGAGTCAGCTGGATGACCAGAGTGCTGAAACCCACAGCCACAAGCAGTCCAAAGTCAGCCGTG AATTCCACAGCCATGAATTTACAGCCATGAAGATATGCTGGTTGTAGACCCCAAAAGTAAGGAAGAAGA TAAACACCTGAAATTTCTGATTTTCTCATGAATTAGATAGTGCATCTTCTGAGGTCAATCTCGAG			
	ORF Start: ATG at 1		ORF Stop: at 834	
	SEQ ID NO: 120		278 aa	MW at 31282.25 Da
NOV11b, 209934449 Protein Sequence	GSIPVKQADSGSSEKQLYNKYPDAVATWLNPDPSQKQNLAPQNAVSSEETNDFKQETL PSKSNESHDMDDMDEDDDDHVDSDSIDSNDSDDVDDTDDSHQSDSHSDESDELVT DFPTDLPADEVFTPVVPTVDTYDGRGDSVVYGLRSKSKKFRRPDIQYPDATDEDITSHME SEELNGAYKAIPVAQDLNAPSDWDSRGKDSYETSQLDDQSAETHSHKQSKVSREFHSHEF HSHEDMLVDPKSKEEDKHLKFRISHELDSASSEVNLE			

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.8200 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 17 and 18

- 5 A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11C.

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB30573	A human Eta-1/osteopontin-a protein - Homo sapiens, 314 aa. [WO200063241-A2, 26-OCT-2000]	1..290 1..314	290/314 (92%) 290/314 (92%)	e-168
AAE12683	Human osteopontin (OPN) - Homo sapiens, 314 aa. [WO200171358-A1, 27-SEP-2001]	1..290 1..314	290/314 (92%) 290/314 (92%)	e-168
AAB01351	Human osteopontin - Homo sapiens,	1..290	290/314 (92%)	e-168

	314 aa. [WO200033865-A1, 15-JUN-2000]	1..314	290/314 (92%)	
AAB19770	Human osteopontin - Homo sapiens, 314 aa. [WO200062065-A1, 19-OCT-2000]	1..290 1..314	290/314 (92%) 290/314 (92%)	e-168
AAW99779	Human osteopontin - Homo sapiens, 314 aa. [WO9908730-A1, 25-FEB-1999]	1..290 1..314	290/314 (92%) 290/314 (92%)	e-168

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

<b>Table 11D. Public BLASTP Results for NOV11a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV11a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P10451	Osteopontin precursor (Bone sialoprotein 1) (Urinary stone protein) (Secreted phosphoprotein 1) (SPP-1) (Nephropontin) (Uropontin) - Homo sapiens (Human), 314 aa.	1..290 1..314	290/314 (92%) 290/314 (92%)	e-167
Q961Z1	Secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1) - Homo sapiens (Human), 300 aa.	1..290 1..300	276/314 (87%) 276/314 (87%)	e-156
CAC16643	Sequence 5 from Patent WO0063241 - Homo sapiens (Human), 287 aa.	1..290 1..287	263/314 (83%) 263/314 (83%)	e-145
P31097	Osteopontin precursor (Bone sialoprotein 1) (Secreted phosphoprotein 1) (SPP-1) (OC-1) - Oryctolagus cuniculus (Rabbit), 311 aa.	1..290 1..311	200/315 (63%) 242/315 (76%)	e-110
P14287	Osteopontin precursor (Bone sialoprotein 1) (Secreted phosphoprotein 1) (SPP-1) - Sus scrofa (Pig), 303 aa.	1..290 1..303	204/309 (66%) 231/309 (74%)	e-104

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.

<b>Table 11E. Domain Analysis of NOV11a</b>			
<b>Pfam Domain</b>	<b>NOV11a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
Osteopontin	1..290	245/334 (73%) 290/334 (87%)	6.7e-198

**Example 12.**

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 49		1042 bp
NOV12a, CG111683-01 DNA Sequence	ATGGATGTGGGCAGCAAAGAGGTCTCTGATGGAGAGCCCGCCGCGTGTCTAGGACTACT CCGCAGCTCCCCGGGGCCGATTGTCATTCCCTGCTGCCAGTGCACCTGAAACGCCT TCTTATCGTGGTGGTGGTGGTGGTCTCCATCGTCTGGTGTATTGTGGGAGCCCTGCTC ATGGGTCTCCACATGAGCCAGAAACACTTTCCCCAGGTTCTGGAGATGAGCATTGGGG CGCCGGAAGCCCAGCAACGCCTGGCCCTGAGTGAGCACCTGGTTACCACTGCCACCTT CTCCATCGGCTCCACTGGCCTCGTGGTGTATGACTACCAGCAGCTGCTGATCGCCTAC AAGCCAGCCCCTGGCACCTGCTGCTACATCATGAAGATAGCTCCAGAGAGCATCCCCA GTCTTGAGGCTCTCACTAGAAAAGTCCACAACCTCCAGGCCAAGCCCGCAGTGCCTAC GTCTAAGCTGGGCCAGGCAGAGGGGCGAGATGCAGGCTCAGCACCTCCGGAGGGGAC CCGGCCTTCCTGGGCATGGCCGTGAGCACCTGTGTGGCGAGGTGCCGCTCTACTACA TCTAGGACGCCTCCGGTGAGCAGGTGTGATCCCAGGGCCCCCTGATCAGCAGCGGAGGA GCGCTCGGGCCACCTGCCCGGGCTGTGGAGGAGCGCTCGCGCTGACCAGGCGCTGGGG CGTCCACTGAAGCGGGGTATCCAGGCAACTCGGGGGAGGGGAAGCTCACAGACCGGT ACTTCCCACCTCCCCTGAATTCTCTCTGTCCATCCTCAACATTCCTTTGCTTCACAGGG TCAGTGGAAGCCCCAACGGGAAAGGAAACGCCCCGGGCAAAGGGTCTTTTGCACTTT TGCAGACGGGCAAGAAGCTGCTTCTGCCACACCGCAGGGACAAACCCTGGAGAAATG GGAGCTTGGGGAGAGGATGGGAGTGGGCAGAGGTGGCACCCAGGGGCCCCGGGAACCTC TGCCACAACAGAATAAAGCAGCCTGATTGAAAAGCAAAAAAAAAAAAAAAAAAACTC		
	ORF Start: ATG at 1		ORF Stop: TAG at 583
	SEQ ID NO: 50		194 aa      MW at 20634.0 Da
NOV12a, CG111683-01 Protein Sequence	MDVGSKEVLMESPPPCQDYSAPRGRFGIPCCPVHLKRLLIIVVVVVSIVVVIIVGALL MGLHMSQKHFPQVLEMSIGAPEAQRLALSEHLVTTATFSIGSTGLVVYDYQQLLIAY KPAPGTCCYIMKIAPESIPSLEALTRKVHNFQAKPAVPTSKLGQAEGRDAGSAPSGGD PAFLGMAVSTLCGEVPLYI		
	SEQ ID NO: 51		590 bp
NOV12b, CG111683-02 DNA Sequence	ATGGATGTGGGCAGCAAAGAGGTCTCTGATGGAGAGCCCGCCGACTACTCCGCAGCTC CCCGGGGCCGATTGTCATTCCCTGCTGCCAGTGCACCTGAAACGCCTTCTTATCGT GGTGGTGGTGGTGGTGGTCTCATCGTCTGGTGTATTGTGGGAGCCCTGCTCATGGGTCTC CACATGAGCCAGAAACACACGGAGATGGTTCTGGAGATGAGCATTGGGGCGCCGGAAG CCCAGCAACGCCTGGCCCTGAGTGAGCACCTGGTTACCACTGCCACCTTCTCCATCGG CTCCACTGGCCTCGTGGTGTATGACTACCAGCAGCTGCTGATCGCCTACAAGCCAGCC CCTGGCACCTGCTGCTACATCATGAAGATAGCTCCAGAGAGCATCCCCAGTCTTGAGG CTCTCAATAGAAAAGTCCACAACCTTCCAGGCCAAGCCCGCAGTGCCTACGTCTAAGCT GGGCCAGGCAGAGGGGCGAGATGCAGGCTCAGCACCTCCGGAGGGGACCCGGCCTTC CTGGGCATGGCCGTGAACACCCTGTGTGGCGAGGTGCCGCTCTACTACATCTAGGCGC CTCCGGTGAG		
	ORF Start: ATG at 1		ORF Stop: TAG at 574
	SEQ ID NO: 52		191 aa      MW at 20360.8 Da
NOV12b, CG111683-02 Protein Sequence	MDVGSKEVLMESPPDYSAPRGRFGIPCCPVHLKRLLIIVVVVVLIVVVIIVGALLMGL HMSQKHTEMVLEMSIGAPEAQRLALSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPA PGTCCYIMKIAPESIPSLEALNRKVHNFQAKPAVPTSKLGQAEGRDAGSAPSGGDPAF LGMAVNTLCGEVPLYI		
	SEQ ID NO: 53		530 bp
NOV12c, CG111683-03	TGGATGTGGGCAGCAAAGAGGTCTCTGATGGAGAGCCCGCCGACTACTCCGCAGCTCC CCGGGGCCGATTGTCATTCCCTGCTGCCAGTGCACCTGAAACGCCTTCTTATCGTG		

DNA Sequence	GTGGTGGTGGTCCTCATCGTCGTGGTGATTGTGGAAGCCCAGCAACGCCTGGCCCTGAGTGAGCACCTGGTTACCACTGCCACCTTCTCCATCGGCTCCACTGGCCTCGTGGTGATGACTACCAGCAGCTGCTGATCGCCTACAAGCCAGCCCCCTGGCACCTGCTGCTACATCATGAAGATAGCTCCAGAGAGCATCCCCAGTCTTGAGGCTCTCAATAGAAAAGTCCACA ACTTCCAGATGGAATGCTCTCTGCAGGCCAAGCCCGCAGTGCCTACGTCTAAGCTGGGCCAGGCAGAGGGGCGAGATGCAGGCTCAGCACCTCCGGAGGGGACCCGGCCTTCCTGGGCATGGCCGTGAACACCCTGTGTGGCGAGGTGCCGCTCTACTACATCTAGGACGCCTCCGGTGAG		
	ORF Start: at 3	ORF Stop: TAG at 513	
	SEQ ID NO: 54	170 aa	MW at 18158.0 Da
NOV12c, CG111683-03 Protein Sequence	DVGSKEVLMEPPDYSAAPRGRFGIPCCPVHLKRLLIVVVVVLIVVVIVEAQQLALS EHLVTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCYIMKIAPEIPSLALNRKVHNFQMECSLQAKPAVPTSKLGQAEGRDAGSAPSGGDPFLGMAVNTLCGEVPLYII		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b and NOV12c.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	1..194	170/194 (87%)
	1..191	171/194 (87%)
NOV12c	2..194	147/199 (73%)
	1..170	148/199 (73%)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 57 and 58

5

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB58144	Lung cancer associated polypeptide sequence SEQ ID 482 - Homo sapiens,	1..194 26..216	187/194 (96%) 187/194 (96%)	e-102

	216 aa. [WO200055180-A2, 21-SEP-2000]			
AAP82978	Human SP5 protein - Homo sapiens, 197 aa. [WO8805820-A, 11-AUG-1988]	1..194 1..197	187/200 (93%) 187/200 (93%)	e-100
AAP70440	Sequence of a canine 5 kd alveolar surfactant protein (ASP) from clone cDNA #19 - Dog, 197 aa. [WO8706588-A, 05-NOV-1987]	1..194 1..197	187/200 (93%) 187/200 (93%)	e-100
AAR15609	SP-5 clone #19 - Homo sapiens, 197 aa. [WO9118015-A, 28-NOV-1991]	1..194 1..197	186/200 (93%) 187/200 (93%)	2e-99
AAP90038	Deduced sequence of cDNA number 19 encoding human SP-5-derived protein - Homo sapiens, 197 aa. [WO8904326-A, 18-MAY-1989]	1..194 1..197	186/200 (93%) 187/200 (93%)	2e-99

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

<b>Table 12E. Public BLASTP Results for NOV12a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV12a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P11686	Pulmonary surfactant-associated protein C precursor (SP-C) (SP5) (Pulmonary surfactant-associated proteolipid SPL(Val)) - Homo sapiens (Human), 197 aa.	1..194 1..197	185/200 (92%) 186/200 (92%)	2e-98
P55152	Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary surfactant-associated proteolipid SPL(Val)) - Macaca mulatta (Rhesus macaque), 191 aa.	1..194 1..191	174/194 (89%) 176/194 (90%)	5e-92
Q9N276	Pulmonary surfactant-associated protein C - Ovis aries (Sheep), 190 aa.	1..193 1..189	159/193 (82%) 169/193 (87%)	9e-83
Q9BDX5	Pulmonary surfactant-associated protein C proSP-C - Bos taurus (Bovine), 190 aa.	1..193 1..189	156/193 (80%) 168/193 (86%)	5e-81
P35245	Pulmonary surfactant-associated protein C precursor (SP-C) - Mustela vison (American mink), 190 aa.	1..194 1..190	154/194 (79%) 167/194 (85%)	1e-80

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PSAP	27..194	150/171 (88%) 164/171 (96%)	6.2e-126

**Example 13.**

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 55	1659 bp	
NOV13a, CG112655-01 DNA Sequence	CGGGCCATGGCCAGAGACCCCCTCCTCTGGGCTCCCTGAAGTCCTGGGGAGCCGTGAC CCATGGGATCGTCGAGCAGCCGGGTGCTGGGCCAGCCGAGGCGAGCCCTTGCCCAGCA GGAACAGGGTGCCAGGGCCAGGGGCTCGGCCCGGAGGCCGACACTGGAGACGATGCG GCGAGCTACGGCTTCTGTTACTGCCCCGGGCAGTCACAAGCGCAAGCGGAGCAGCGGGG CCTGCCGCTACTGTGACCCGGACTCGCACAGGGAGGAGCATGAGGAGGAGGGGGACAA GCAGCAGCCGCTCCTCAACACCCCTGCAAGGAAAAAATTAAGGAGTACATCCAAATAT ATTTATCAAACATTATTTTTGAATGGTGAAAACAGTGACATTAAGATTTGTGCTCTAG GAGAAGAATGGCGATTACACAAAATATATTTATGTCAATCTGGCTACTTTTCTAGTAT GTTTCAGTGGTTCCTTGGAAAGAATCCAGCATGAATATTATTGAACTGGAGATTCCTGAC CAGAACATTGATGTAGACGCACTGCAGGTTGCGTTTGGTTCACTGTATCGAGATGATG TCTTGATAAAACCCAGTCGAGTTGTTGCCATTTTGGCAGCAGCTTGATGCTGCAGCT GGATGGTTTAAATACAGCAGTGTGGTGAGACAATGAAGGAAACAATTAATGTGAAAAC GTATGCGGTTATTACACATCAGTAGAGATCTATGGATTAGATTCTGTAAAGAAAAAGT GCCTTGAATGGCTTCTAAACAATTTGATGACTCACCAGAATGTTAACTTTTTAAAGA ACTCGGTATAAATGTCATGAAACAGCTCATTGGTTCCTCTAACTTATTTGTGATGCAA GTGGAGATGGATGTATACACCACTCTAAAAAAGTGGATGTTCCTTCAACTTGTGCCTT CTTGGAATGGATCTTTAAAAACAGCTTTTGACAGAAACAGATGTCTGGTTTCTTAAACA GAGAAAAGATTTTGAAGGTATGGCCTTTCTTGAACTGAACCAGGAAAACCATTTGTG TCAGTATTCAGACATTTAAGGTTACAATATATTATCAGTGACCTAGCTTCTGCAAGAA TTATTGAACAAGATGGTATAGTACCTTCAGAATGGCTGTCTTCTGTGTATAAACAGCA GTGGTTTGCTATGCTGCGGGCAGAACAAAGACCATGAGGTAGGGCCTCAAGAAATCAAT AAAGAAGACCTAGAGGGAAGTAGCATGAGGTGTGGTAGAAAGCTTGCCAAAGATGGTG AATACTACTGGTGTGGACGGGTTTTAACTTCGGCTTTGACCTACTTGTAATTTACAC CAATGGATACATCATTTTCAAACGCAATACACTGAATCAGCCATGCAGCGGGTCTGTC AGTTTACGGCCTCGAAGGAGCATAGCATTTAGATTACGCTTGGCTTCTTTTGATAGTA GTGGAAAAC TAGTATGTAGTAGAACA ACTGGCTATCAAATACTTATACTTAAAAAGGA TCAGGAACAAGTGGTGATGAACTTGACAGCAGGTTTCTGACCTTCCCTTTATATATC TGCTGTAACCTTCTTGATATATCACCAGAAAAGGAATTGAAAATAATCGCCACCCAG AAGATCCAGAAAAC <u>TGAAGATCTCATCAGTTGGAA</u>		
	ORF Start: ATG at 61	ORF Stop: TGA at 1639	
	SEQ ID NO: 56	526 aa	MW at 60200.2 Da
NOV13a, CG112655-01 Protein Sequence	MGSSSSRVLGQPRRALAQEQGARARGSSARRPDTGDDAASYGFCYCPGSHKRKRSSGA CRYCDPDSHREEHEEEGDKQQLLNT PARKKLRSTSKYIYQTLFLNGENSDIKICALG EEWRLHKIYLCQSGYFSSMFGSGWKESMNI I E L E I P D Q N I D V D A L Q V A F G S L Y R D D V L I K P S R V V A I L A A C M L Q L D G L I Q C G E T M K E T I N V K T V C G Y Y T S V E I Y G L D S V K K K C L E W L L N L M T H Q N V K L F K E L G I N V M K Q L I G S S N L F V M Q V E M D V Y T T L K K W M F L Q L V P S W N G S L K Q L L T E T D V W F S K Q R K D F E G M A F L E T E P G K P F V S V F R H L R L Q Y I I S D L A S A R I I E Q D G I V P S E W L S S Y K Q Q W F A M L R A E Q D H E V G P Q E I N K E D L E G S S M R C G R K L A K D G E Y Y W C W T G F N F G F D L L V I Y T N G Y I I F K R N T L N Q P C S G S V S L R P R R S I A F R L R L A S F D S S		

	GKLVCSTRTGYQILILKKDQEQVVMNLDSTRFLTPLYICCNFLYISPEKGIENNRHPE DPEN
--	--------------------------------------------------------------------

Further analysis of the NOV13a protein yielded the following properties shown in Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.6850 probability located in plasma membrane; 0.4605 probability located in mitochondrial inner membrane; 0.3500 probability located in nucleus; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13C.

Table 13C. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB94442	Human protein sequence SEQ ID NO:15072 - Homo sapiens, 515 aa. [EP1074617-A2, 07-FEB-2001]	1..513 1..513	465/513 (90%) 482/513 (93%)	0.0
AAB95625	Human protein sequence SEQ ID NO:18346 - Homo sapiens, 510 aa. [EP1074617-A2, 07-FEB-2001]	1..510 1..510	462/510 (90%) 477/510 (92%)	0.0
AAY18025	Murine DIP protein sequence - Mus sp, 524 aa. [WO9927091-A1, 03-JUN-1999]	1..524 1..522	442/524 (84%) 470/524 (89%)	0.0
AAY01080	Human testis specific growth factor, ZGCL-1, protein sequence - Homo sapiens, 478 aa. [WO9909168-A1, 25-FEB-1999]	48..513 12..477	427/466 (91%) 442/466 (94%)	0.0
AAB94515	Human protein sequence SEQ ID NO:15231 - Homo sapiens, 381 aa. [EP1074617-A2, 07-FEB-2001]	135..513 1..379	352/379 (92%) 362/379 (94%)	0.0

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13D.

Table 13D. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q8TC88	Hypothetical 60.2 kDa protein - Homo sapiens (Human), 526 aa.	1..526 1..526	525/526 (99%) 526/526 (99%)	0.0
Q8TC89	Hypothetical 60.2 kDa protein - Homo sapiens (Human), 526 aa.	1..526 1..526	524/526 (99%) 525/526 (99%)	0.0
Q96IK5	Hypothetical 58.7 kDa protein - Homo sapiens (Human), 515 aa.	1..513 1..513	466/513 (90%) 482/513 (93%)	0.0
Q9H927	CDNA FLJ13057 fis, clone NT2RP3001580, highly similar to Mus musculus strain C57BL/J germ cell-less protein (Gcl) mRNA - Homo sapiens (Human), 515 aa.	1..513 1..513	465/513 (90%) 482/513 (93%)	0.0
Q9H826	CDNA FLJ13980 fis, clone Y79AA1001692, weakly similar to germ cell-LESS protein - Homo sapiens (Human), 511 aa (fragment).	1..511 1..511	463/511 (90%) 478/511 (92%)	0.0

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value
BTB	92..208	23/144 (16%) 83/144 (58%)	1.5e-11

#### Example 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis		
	SEQ ID NO: 57	1225 bp
NOV14a, CG112813-01 DNA Sequence	TGGCACCATGGCCCCAAACTCATCACCGTCCTGTGTCTGGGATTCTGCCTGAACCAG AAGATCTGCCCACATGCGGGTGCTCAGGACAAGTTCTCCCTGTCAGCCTGGCCGAGCC CTGTGGTTCCCCCTAGGAGGACGTGTGACTCTCTCCTGTCAATCCCATCTTCGGTTTGT CATATGGACAATATTCCAACAACCTGGGACCCGAAGCCATGAGTTGCACACTGGCCTT TCCAACAACATCACCATCAGCCCTGTGACCCCAAGACACGCAGGGACCTACAGATGTG TTGGAATTTACAAGCACGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTGAAGATCAT CGTCACAGGCTTGTTCAAAAACCTCCATCTCAGCGCACCCAAGCTCCCTGGTGCAT GCAGGAGCCAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGAATTTATCT TATACAAAGAGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATGGAGGCTGG GATCCATTACGTCGAGGCTGTCTTTTCCATGGGTCTGTAACGCCTGCCCATGCAGGA GCCTACAGATGCTGTGGTTGTTTCAGTCACTCCCGCTATGAGTGGTCGGCTCCCAGTG ACCCCTGGACATTGTGATCACAGGAAAATACAAAAGCCTTCTCTCCACCCAGGT GGACCCCATGATGAGGCTGGGAGAGAAGTTGACCCCTCTTCTGCAGCTCTGAAATCTCA TTGACCACTACCATCTGTTTCAGACACGGGGTTGCTCATGGACAGTGGCTCAGTGGAG GGCAGAGACACAGGGAAGCATTCCAGGCCAATTTTCTGTGGGGCCGTGCAACGCCAGT CCCTGGCGGGACCTATAGATGCTATGGTTCTTCAATGACTCTCCCTATAAGCCCCCA	



	GTGACCCGCTGCAACTTTACACCACAGGAAACACTAAGAGTACTCCTCTGTTCATTAC AGAATCCACCCCTGAATCTGGAGCCTGCAGCAGAAGAGACACAGGAGATCATATATGC CCAGTTAAACCACCAGGCCCTCTCACAGACAGGATTCCCTCCTGCCTCCCAGTGTCCC CACTACCTCTCGGAGGATCCTAGTATCTACATCACTGTCCACCAAGCCCAGGCTGAGG CCAGAGCTGCCCCAGTCTTTGGCACAAAGGGCATTAAATACGCAAGGACCTGGATCTA TTCCTAG		
	ORF Start: ATG at 8	ORF Stop: TAA at 1196	
	SEQ ID NO: 58	396 aa	MW at 43739.2 Da
NOV14a, CG112813-01 Protein Sequence	MAPKLITVLCGLFCLNQKICPHAGAQDKFSLSAWPSPVVPLGGRVTLSCSHSLRFVIW TIFQTTGTRSHLHTGLSNNITISPVTPHAGTYRCVGIYKHASKWSAESNSLKIIVT GLFTKPSISAHPSLVHAGARVSLRCHSELAFDEFILYKEGHIQHSQQLDQGM EAGIH YVEAVFSMGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKKPSLSTQVDP MMRLGEKLTFLCSSEISFDQYHLFRHGVAHQWLSSGGQRHREAFQANFSVGRATPVPG GTYRCYGSFNDSPIKPPVTRCNFTPQETLRVLLCHSQNPPLNLEPAAEETQEIIYAQL NHQALSQTGFPPASQCPHYLSEDPSIYITVHQQAQAEARAAPSLWHKGH		
	SEQ ID NO: 59	1399 bp	
NOV14b, CG112813-02 DNA Sequence	TGGCACCATGGCCCCAAACTCATCACCGTCCTGTGTCTGGGATTCTGCCTGAACCAG AAGATCTGCCACATGCGGGTGCTCAGGACAAGTTCTCCCTGTCAGCCTGGCCGAGCC CTGTGGTTCCCTTAGGAGGACGTGTGACTCTCTCCTGTCAATCCCATCTTCGGTTTGT CATATGGACAATATTCCAAACAACCTGGGACCCGAAGCCATGAGTTGCACACTGGCCTT TCCAACAACATCACCATCAGCCCTGTGACCCAGAACACGCAGGGACCTACAGATGTG TTGGAATTTACAAGCACGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTGAAGATCAT CGTCACAGGCTTGTTCAAAAACCCCTCATCTCAGCGCACCCAAGCTCCCTGGTGAT GCAGGAGCCAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGAATTTATCT TATACAAAGAGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATGGAGGCTGG GATCCATTACGTCGAGGCTGTCTTTTCCATGGGTCTGTAAACGCCTGCCCATGCAGGA GCCTACAGATGCTGTGGTTGTTTCAGTCACTCCCGCTATGAGTGGTCCGCTCCAGTG ACCCCTGGACATTGTGATCACAGGAAAATACAAAAGCCTTCTCTCTCCACCCAGGT GGACCCCATGATGAGGCTGGGAGAGAAGTTGACCCCTCTCTGCAGCTCTGAAATCTCA TTTGACCAGTACCATCTGTTTCAGACACGGGGTTGCTCATGGACAGTGGCTCAGTGGAG GGCAGAGACACAGGGAAGCATTCCAGGCCAATTTTTCTGTGGGCGGTGCAACGCCAGT CCCTGGCGGGACCTATAGATGCTATGGTTCCCTTCAATGACTCTCCCTATAAGCCCCCA GTGACCCACTGCAACTTTACACCACAGGAAACACTAAGAGTACTCCTCTGTCAATTCAC AGAATCCACCCCTGAATCTGACACACCTCGCCCTCAAGGACAGTCCAGCAACCTGCAT ATGCTCACTGGACTCTCAGTAGCCATCATCTCCATTGGCGTTTGCCTCTCTGCTTTTA TTGGTTTCTGGTGTTACATAAAATATCACACCACCATGGCAAACACAGAGCCCACGGA AGGCCAACGGACGGATGAAGAGGAGCCTGCAGCAGAAGAGACACAGGAGATCATATAT GCCCAGTTAAACCACCAGGCCCTCTCACAGACAGGATTCCCTCCTGCCTCCCAGTGTG CCCCTACCTCTCGAAGGATCCTAGTATCTACATCACTGTCCACCAAGCCCAGGCTGA GGCCAGAGCTGCCCCAGTCTTTGGCACAAAGGGCATTAAATACGCAAGGACCTGGATC TATTCCT		
	ORF Start: ATG at 8	ORF Stop: TAG at 1064	
	SEQ ID NO: 60	352 aa	MW at 38757.9 Da
NOV14b, CG112813-02 Protein Sequence	MAPKLITVLCGLFCLNQKICPHAGAQDKFSLSAWPSPVVPLGGRVTLSCSHSLRFVIW TIFQTTGTRSHLHTGLSNNITISPVTPHAGTYRCVGIYKHASKWSAESNSLKIIVT GLFTKPSISAHPSLVHAGARVSLRCHSELAFDEFILYKEGHIQHSQQLDQGM EAGIH YVEAVFSMGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKKPSLSTQVDP MMRLGEKLTFLCSSEISFDQYHLFRHGVAHQWLSSGGQRHREAFQANFSVGRATPVPG GTYRCYGSFNDSPIKPPVTHCNFTPQETLRVLLCHSQNPPLNLTHLALKDSPATCICS LDSQ		
	SEQ ID NO: 61	1369 bp	
NOV14c, CG112813-04 DNA Sequence	ATGGCCCCCAAACCTCATCACCGTCCTGTGTCTGGGATTCTGCCTGAACCAGAAGATCT GCCACATGCGGGTGCTCAGGACAAGTTCTCCCTGTCAGCCTGGCCGAGCCCTGTGGT TCCCTTAGGAGGACGTGTGACTCTCTCCTGTCAATCCCATCTTCGGTTTGTTCATATGG ACAATATTCCAAACAACCTGGGACCCGAAGCCATGAGTTGCACACTGGCCTTTCCAACA		

	ACATCACCATCAGCCCTGTGACCCCAAGACACGCAGGGACCTACAGATGTGTTGGAAT TTACAAGCACGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTGAAGATCATCGTCACA GGCTTGTTTACAAAACCCCTCCATCTCAGCGCACCCAAGCTCCCTGGTGCATGCAGGAG CCAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGAATTTATCTTATACAA AGAGGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATGGAGGCTGGGATCCAC TACGTCGAGGCTGTCTTTTCCATGGGTCTGTAAACGCCTGCCCATGCAGGAGCCTACA GATGCTGTGGTTGTTTCAGTCACTCCCCTATGAGTGGTCGGCTCCCAGTGACCCCT GGACATTGTGATCACAGGAAAATACAAAAGCCTTCTCTCTCCACCCAGGTGGACCCC ATGATGAGGCTGGGAGAGAAGTTGACCCTCTTCTGCAGCTCTGAAATCTCATTTGACC AGTACCATCTGTTTCAGACACGGGGTTGCTCATGGACAGTGGCTCAGTGGAGGGCAGAG ACACAGGGAAGCATTCAGGCCAATTTTTCTGTGGGCCGTGCAACGCCAGTCCCTGGC GGGACCTATAGATGCTATGGTTCTTCAATGACTCTCCCTATAAGCCCCCAGTGACCC ACTGCAACTTTACACCACAGGAAACACTAAGAGTACTCCTCTGTTCATTACAGAATCC ACCCCTGAATCTGACACACCTCGCCCTCAAGGACAGTCCAGCAACCTGCATATGCTCA CTGGACTCTCAGTAGCCATCATCTCCATTGGCGTTTGCCCTCTCTGCTTTTATTGGTTT CTGGTGTTACATAAAATATCACACCACCATGGCAAACACAGAGCCCACGGAAGGCCAA CGGACGGATGAAGAGGAGCCTGCAGCAGAAGAGACACAGGAGATCATATATGCCAGT TAAACCACCAGGCCCTCTCACAGACAGGATTCCCTCCTGCCTCCCAGTGTCCCCACTA CCTCTCGAAGGATCCTAGTATCTACATCACTGTCCACCAAGCCCAGGCTGAGGCCAGA GCTGCCCCCAGTCTTTGGCACAAAGGGCATTATA		
	ORF Start: ATG at 1	ORF Stop: TAG at 1057	
	SEQ ID NO: 62	352 aa	MW at 38757.9 Da
NOV14c, CG112813-04 Protein Sequence	MAPKLITVLCGLFCLNQKICPHAGAQDKFSLSAWPSVVPVPLGGRVTLSCSHSLRFVIW TIFQTTGTRSHLHTGLSNNITISPVTPHEAGTYRCVGIYKHASKWSAESNSLKIIVT GLFTKPSISAHPSLHVHAGARVSLRCHSELAFDEFILYKEGHIQHSQQLDQGM EAGIH YVEAVFSMGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKKPSLSLQVDP MMRLGEKLTLCFCSSEISFDQYHLFRHGVAHQWLSGGQRHREAFQANFSVGRATPVPG GTYRCYGSFNDSFYKPPVTHCNFTPQETLRVLLCHSQNPPLNLTHLALKDSPATCICS LDSQ		
	SEQ ID NO: 63	1502 bp	
NOV14d, CG112813-05 DNA Sequence	ATGGCCCCCAAACCTCATCACCGTCTGTGCTAGGATTCTGCCTGAACCAGAAGATCT GCCCATATGCGGGTGCTCAGGACAAGTTCTCCCTGTCAGCCTGGCCGAGCCCTGTGGT TCCCTAGGAGGACGTGTGACTCTCTCCTGTTCATTCCCATCTTCGGTTTGTCATATGG ACAATATTCCAAACAACCTGGGACCCGAAGCCATGAGTTGCACACTGGCCTTTCCAACA ACATCACCATCAGCCCTGTGACCCCAAGACACGCAGGGACCTACAGATGTGTTGGAAT TTACAAGCACGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTGAAGATCATCGTCACA GGCTTGTTTACAAAACCCCTCCATCTCAGCGCACCCAAGCTCCCTGGTGCATGCAGGAG CCAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGAATTTATCTTATACAA AGAGGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATGGAGGCTGGGATCCAT TACGTCGAGGCTGTCTTTTCCATGGGTCTGTAAACGCCTGCCCATGCAGGAGCCTACA GATGCTGTGGTTGTTTCAGTCACTCCCCTATGAGTGGTCGGCTCCCAGTGACCCCT GGACATTGTGATCACAGGAAAATACAAAAGCCTTCTCTCTCCACCCAGGTGGACCCC ATGATGAGGCTGGGAGAGAAGTTGACCCTCTTCTGCAGCTCTGAAATCTCATTTGACC AGTACCATCTGTTTCAGACACGGGGTTGCTCATGGACAGTGGCTCAGTGGAGGGCAGAG ACACAGGGAAGCATTCAGGCCAATTTTTCTGTGGGCCGTGCAACGCCAGTCCCTGGC GGGACCTATAGATGCTATGGTTCTTCAATGACTCTCCCTATAAGCCCCCAGTGACCC GCTGCAACTTTACACCACAGGAAACACTAAGAGTACTCCTCTGTTCATTACAGAATCC ACCCCTGAATCTGACACCACCATGGCAAACACAGAGCCCACGGAAGGCCAACGGACGG ATGAAGAGGAGCCTGCAGCAGAAGAGACACAGGAGATCATATATGCCAGTTAAACCA CCAGGCCCTCTCACAGACAGGATTCCCTCCTGCCTCCCAGTGTCCCCACTACCTCTCG GAGGATCCTAGTATCTACATCACTGTCCACCAAGCCAGGCTGAGGCCAGAGCTGCC CCAGTCTTTGGCACAAAGGGCATTAAACGCAAGGACCTGGATCTATTCTAGGAGGA TTTTTTTTTCCACGGACATTCTTCTCCTTCTGGTACCATCTTGACACCTCGAAGCTGG CAACAGCAGTGTCTGAATGCTTGTGGGATTATCTTAAATTCAGCACTGCTGAACAG ACAACCTAGCCATTCTACAATTCTATTTTGAGCATCCAACCATTTTCAGGTGATTGACT CTTACCACACACTCATCTGGATATCTCATTAAATATCATCTGAATTATCCTG		
	ORF Start: ATG at 1	ORF Stop: TAA at 1096	

	SEQ ID NO: 64	365 aa	MW at 40669.1 Da
NOV14d, CG112813-05 Protein Sequence	MAPKLITVLC LGFCLNQKICPHAGA QDKFSLSAWPSPVVPLGGRVTL SCHSHLRFVIW TIFQTTGTRSHELHTGLSNNITISPVTP EHAGTYRCVGIYKHASKWSAESNSLKIIVT GLFTKPSISAH PSSLVHAGARVSLRCHSELAFDEFILYKEGHIQHSQQLDQGM EAGIH YVEAVFSMGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKPSLSTQVDP MMRLGEKLT LFCSS EISFDQYHLFRHGVAHGQWLSGGQRHREAFQANFSVGRATPVPG GTYRCYGSFNDS PYKPPVTRCNFTPQETLRVLLCHSQNPPLNLTPPWQTQSPRKANGR MKRSLQQKRHRRSYMPS		
	SEQ ID NO: 65	1327 bp	
NOV14e, CG112813-06 DNA Sequence	AATAGAAGTGGCACCATGGCCCCAAACTCATCACCGTCCTGTGCCTAGGATTCTGCC TGAACCAGAAGATCTGCCACATGCGGGTGCTCAGGACAAGTTCTCCCTGTCAGCCTG GCCGAGCCCTGTGGTTCCCCTAGGAGGACGTGTGACTCTCTCCTGTCATTCCCCTCTT CGGTTTGT CATATGGACAATATTC AAACA ACTGGGACCCGAAGCCATGAGTTGCACA CTGGCCTTTCCAACAACATCACCATCAGCCCTGTGACCCGACACGACGGGACCTA CAGATGTGTGGAATTTACAAGCAGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTG AAGATCATCGTCACAGGTAGGTTCAAAAACCCTCCATCTCAGCGCACCCAAGCTCCC TGGTGCATGCAGGAGCCAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGA ATTTATCTTATACAAAGAGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATG GAGGCTGGGATCCATTACGTCGAGGCTGTCTTTTCATGGGTCCTGTAACGCCTGCCC ATGCAGGAGCCTACAGATGCTGTGGTTGTTTTAGTCACTCCCGCTATGAGTGGTCGGC TCCCAGTGACCCCTGGACATTGTGATCACAGGTAAATACAAAAGCCTTCTCTCTCC ACCCAGGTGGACCCCATGATGAGGCTGGGAGAGAAGTTGACCCTCTTCTGCAGCTCTG AAATCTCATTTGACCA GTACCATCTGTTTCAGACACGGGGTTGCTCATGAGATGGCT CAGTGAGGGCAGAGACACAGGGAAGCATTCCAGGCCAATTTTTCTGTGGGCCGTGCA ACGCCAGTCCCTGGCGGGACCTATAGATGCTATGGTTCCTTCAATGACTCTCCCTATA AGACAGACACACCTCGCCCTCAAGGACAGTCCAGCAACCTGCATATGCTCACTGGACT CTCAGTAGCCATCATCTCCATTGGCGTTTGCCTCTCTGCTTTTATTGGTTTCTGGTGT TACATAAAATATCACACCACCATGGCAAACACAGAGCCCACGGAAGGCCAACGGACGG ATGAAGAGGAGCCTGCAGCAGAAGAGACACAGGAGATCATATATGCCCAGTTAAACCA CCAGGCCCTCTCACAGACAGGATTCCCTCCTGCCTCCAGTGTCCCCACTACCTCTCG AAGGATCCTAGTATCTACATCACTGTCCACCAAGCCCAGGCTGAGGCCAGAGCTGCC CCAGTCTTTGGCACAAGGGCATTAAATACGCAAGGACCTGGATCTATTCTT		
	ORF Start: ATG at 16	ORF Stop: TAA at 1300	
	SEQ ID NO: 66	428 aa	MW at 47211.0 Da
NOV14e, CG112813-06 Protein Sequence	MAPKLITVLC LGFCLNQKICPHAGA QDKFSLSAWPSPVVPLGGRVTL SCHSHLRFVIW TIFQTTGTRSHELHTGLSNNITISPVTP EHAGTYRCVGIYKHASKWSAESNSLKIIVT GRFTKPSISAH PSSLVHAGARVSLRCHSELAFDEFILYKEGHIQHSQQLDQGM EAGIH YVEAVFSMGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKPSLSTQVDP MMRLGEKLT LFCSS EISFDQYHLFRHGVAHGQWLSGGQRHREAFQANFSVGRATPVPG GTYRCYGSFNDS PYKTDTPRPQGQSSNLHMLTGLSVAIISIGVCLSAFIGFWCYIKYH TTMANTEPTEGQRTDEEPPAAEETQEIIYAQLNHQALSQTGFPPASQCPHYLSKDPSI YITVHQAQAEARAAPSLWHKGH		
	SEQ ID NO: 67	780 bp	
NOV14f, 209886463 DNA Sequence	AAGCTTGGAGGACGTGTGACTCTCTCCTGTGTCATTCCCCTCTCGGTTTGT CATATGGA CAATATTC AAACA ACTGGGACCCGAAGCCATGAGTTGCACACTGGCCTTTCCAACA CATCACCATCAGCCCTGTGACCCGAGAACACGCAGGGACCTACAGATGTGTTGGAATT TACAAGCACGCCTCAAAGTGGTCAGCTGAGAGCAACTCCCTGAAGATCATCGTCACAG GCTTGTTCAAAAACCCTCCATCTCAGCGCACCCAAGCTCCCTGGTGCATGCAGGAGC CAGGGTGAGCCTGCGCTGTCACTCAGAACTGGCCTTTGATGAATTTATCTTATACAAA GAGGGGCACATACAGCATTCCCAGCAGCTTGACCAGGGGATGGAGGCTGGGATCCATT ACGTCGAGGCTGTCTTTTCATGGGTCCTGTAACGCCTGCCCATGTAGGAGCCTACAG ATGCTGTGGTTGTTTCAGTCACTCCCGCTATGAGTGGTCGGCTCCCACTGACCCCTG GACATTGTGATCACAGGAAAATACAAAAGCCTTCTCTCTCCACCAGGTGGACCCCA TGATGAGGCTGGGAGAGAAGTTGACCCTCTTCTGCAGCTTGAAATCTCATTTGACCA GTACCATCTGTTTCAGACACGGGGTTGCTCATGGACAGTGGCTCAGTGGAGGGCAGAGA CACAGGGAAGCATTCCAGGCCA ACTTTTTCTGTGGGCCGTGCAACGCCAGTCCCTGGCG		

	GGACCTATAGATGCTATGGTCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 68	260 aa	MW at 28816.5 Da
NOV14f, 209886463 Protein Sequence	KLGGRVTLSCSHSLRFVIWTIFQTTGTRSHELHTGLSNNITISPVTPPEHAGTYRCVGI YKHASKWSAESNSLKIIVTGLFTKPSISAHPSLVHAGARVSLRCHSELADEFILYK EGHIQHSQQLDQGM EAGIHYVEAVFSMGPVTPAHVGAYRCCGCFSHSRYEWSAPSDPL DIVITGKYKKPSLSTQVDPMMRLGEKLTFLCSSEISFDQYHLFRHGV AHGQWLSGGQR HREAFQANFSVGRATPVPGGTYRCYGLE		
	SEQ ID NO: 69	871 bp	
NOV14g, 277731421 DNA Sequence	GCCAAGCTTCATGAGTTGCACACTGGCCTTTCCAACAACATCACCATCAGCCCTGTGA CCCCAGAACACGCAGGGACCTACAGATGTGTTGGAATTTACAAGCACGCCTCAAAGTG GTCAGCTGAGAGCAACTCCCTGAAGATCATCGTCACAGGCTTGTTCAAAAACCCCTCC ATCTCAGCGCACCCAAGCTCCCTGGTGCATGCAGGAGCCAGGGTGAGCCTGCGCTGTC ACTCAGAACTGGCCTTTGATGAATTTATCTTATACAAAGAGGGGCACATACAGCATTTC CCAGCAGCTTGACCAGGGGATGGAGGCTGGGATCCACTACGTCGAGGCTGTCTTTTCC ATGGGTCCCTGTAACGCCTGCCCATGCAGGAGCCTACAGATGCTGTGGTTGTTTCAGTC ACTCCCGCTATGAGTGGTCGGCTCCCAGTGACCCCTGGACATTGTGATCACAGGAAA ATACAAAAGCCTTCTCTCTCCACCCAGGTGACCCCATGATGAGGCTGGGAGAGAAG TTGACCCTCTTCTGCAGCTCTGAAATCTCATTGACCAGTACCATCTGTTTCAGACACG GGGTTGCTCATGGACAGTGGCTCAGTGGAGGGCAGAGACACAGGGAAGCATTCCAGGC CAATTTTTCTGTGGGCCGTGCAACGCCAGTCCCTGGCGGGACCTATAGATGCTATGGT TCCTTCAATGACTCTCCCTATAAGCCCCCAGTGACCCACTGCAACTTTACACCACAGG AAACACTAAGAGTACTCCTCTGTCTATTACAGAATCCACCCCTGAATCTGACACACCT CGCCCTCAAGGACAGTCCAGCAACCTGCATATGCTCACTGGACTCTCAGCTCGAGGGT G		
	ORF Start: at 1	ORF Stop: at 871	
	SEQ ID NO: 70	290 aa	MW at 31948.9 Da
NOV14g, 277731421 Protein Sequence	AKLHELHTGLSNNITISPVTPPEHAGTYRCVGIYKHASKWSAESNSLKIIVTGLFTKPS ISAHPSLVHAGARVSLRCHSELADEFILYKEGHIOHSQQLDQGM EAGIHYVEAVFS MGPVTPAHAGAYRCCGCFSHSRYEWSAPSDPLDIVITGKYKKPSLSTQVDPMMRLGEK LTLFCSSEISFDQYHLFRHGV AHGQWLSGGQRHREAFQANFSVGRATPVPGGTYRCYG SFNDSPYKPPVTHCNFTPQETLRVLLCHSQNPPLNLTHLALKD SPATCICSLDSQLEG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 14B.

Table 14B. Comparison of NOV14a against NOV14b through NOV14g.		
Protein Sequence	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV14b	1..333	332/333 (99%)
	1..333	332/333 (99%)
NOV14c	1..333	332/333 (99%)
	1..333	332/333 (99%)
NOV14d	1..335	334/335 (99%)
	1..335	334/335 (99%)
NOV14e	1..396	366/428 (85%)
	1..428	370/428 (85%)
NOV14f	41..297	256/257 (99%)

	2..258	256/257 (99%)
NOV14g	69..333	264/265 (99%)
	4..268	264/265 (99%)

Further analysis of the NOV14a protein yielded the following properties shown in Table 14C.

Table 14C. Protein Sequence Properties NOV14a	
PSort analysis:	0.4489 probability located in lysosome (lumen); 0.3700 probability located in outside; 0.2307 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 69 and 70

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14D.

5

Table 14D. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG10169	Novel human diagnostic protein #10160 - Homo sapiens, 444 aa. [WO200175067-A2, 11-OCT-2001]	1..305 1..303	145/307 (47%) 190/307 (61%)	5e-71
ABG10165	Novel human diagnostic protein #10156 - Homo sapiens, 491 aa. [WO200175067-A2, 11-OCT-2001]	1..386 65..486	165/426 (38%) 228/426 (52%)	5e-71
AAM25638	Human protein sequence SEQ ID NO:1153 - Homo sapiens, 444 aa. [WO200153455-A2, 26-JUL-2001]	1..305 1..303	145/307 (47%) 190/307 (61%)	5e-71
ABG10169	Novel human diagnostic protein #10160 - Homo sapiens, 444 aa. [WO200175067-A2, 11-OCT-2001]	1..305 1..303	145/307 (47%) 190/307 (61%)	5e-71
ABG10167	Novel human diagnostic protein #10158 - Homo sapiens, 388 aa. [WO200175067-A2, 11-OCT-2001]	1..305 1..303	142/307 (46%) 191/307 (61%)	7e-70

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14E.

Table 14E. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H7L2	FLJ00060 protein - Homo sapiens (Human), 227 aa (fragment).	114..333 5..224	217/220 (98%) 220/220 (99%)	e-131
Q99563	NK receptor - Homo sapiens (Human), 436 aa.	1..382 1..435	171/439 (38%) 228/439 (50%)	1e-71
AAK30061	Killer cell immunoglobulin-like receptor 3DL1 - Homo sapiens (Human), 444 aa.	5..305 5..303	144/303 (47%) 191/303 (62%)	3e-71
Q9UER1	KIR3DL1-like natural killer cell receptor - Homo sapiens (Human), 444 aa.	5..305 5..303	144/303 (47%) 191/303 (62%)	3e-71
AAF61292	Killer cell immunoglobulin receptor variant - Homo sapiens (Human), 444 aa.	5..305 5..303	143/303 (47%) 190/303 (62%)	3e-70

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14F.

Table 14F. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig	42..96	17/59 (29%) 42/59 (71%)	5e-07
ig	135..197	11/67 (16%) 44/67 (66%)	0.00019
ig	237..297	14/65 (22%) 42/65 (65%)	0.0018

#### Example 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 71	4380 bp
NOV15a, CG112869-01 DNA Sequence	ATATCTGTGGATGCTATGCATGTCTTCATTGATGAACATGGTGAGGGGGAAATTAGAT CCTGTTATTTAAAATCTGGAAATCAGAAAGAAGGCCCTTTACAGCCTCTACCATCAAA TAATGACTGTCTCTCTCAGGCTAGAGAGATGCAGGTCAGCTCCTCCAGTACCACAAC TCTGAGAGTCAAGATCCGTCTTCTGTTGGGACCCCTGCCGTGAGTGCCTTCAGCAACAGC TGTACTGATGGTGGCTCGCAGGACCCAGTCGGAAACCCACGGCATGTGAGTCAGGA	

TCTGGAAGCCTCGTCATGTTCTTCAACACAAGGAAAATTTAACCGAGAGCAGTTTTAC  
 AAATTTATCATTTTTCCCTGGCAAGTGGATTAAAGTCTGGTATGATCGACTGACCTTGC  
 TGGCATTACTTGATCGGACTGAAGACATCAAGGAGAATGTACTGGCGATTTTTACTCAT  
 TGTCCTGGTTTTCCCTCCTTGGATTTCTGACCTTGAGCCAAGGCTTTTGCAAAGATATG  
 TGGGTGCTCCTCTTCTGCCTCGTCATGGCCAGCTGCCAGTACTCCCTGCTAAAGAGTG  
 TTCAGCCTGACCCCGCTCACCAATACACGGACACAACCAAATCATAACATATAGCAG  
 ACCAATCTATTTTTGTGTGCTGTGTGGCCTTATTTTGCTTCTTGATACAGGGGCCAAA  
 GCCAGGCACCCTCCCAGTTACGTTGTGTATGGCCTGAAGCTCTTCTCCAGTGTTC  
 TACAATCAGCTAGGGACTACTTAATAGTATTTTTATATTGCTTCCCTGCTATTTCCCT  
 CCTTGGGCTCTTCCCGCAAATCAACACTTTCTGCACCTTATCTTTGGGACAAATTGAC  
 ATGCTGTTTTTTGGTGGTTCTGCTGTGTCTGGGATAACCTCGGCTGTTTACAGTGTGG  
 CCCGGAGCGTCTTGGCTGCCGCCCTGCTCCACGCAGTCTGCTTCAGTGCAGTGAAGGA  
 ACCGTGGAGCATGCAACACATCCCGGCACTGTTTTCGGCCTTCTGTGGCCTCTTGGTC  
 GCCCTTTCTTACCATCTGAGCCGTCAGAGCAGTGACCCATCTGTACTCTTTTCCACTT  
 TCAGGTCCTTCATCCAATGCAGGCTGTTTCCATAATTTTTACATCAAAATCTGGCAGA  
 GTCAGCTGCTGACCCTCTCCCCAAGAAGATGAAAGATTTCAGTGGTGAGACATTTGCGT  
 TTTAAATGGGATCTCATCGTCTGCGCAGTGGTTGCTGTCCTCTCATTGTGCAGTCAGCG  
 CCAGACTGTATTCTGTGTCATTGCAGCCATTCTCAGCATCGTGCTGTTTGCCTTGGC  
 TGGAGCCGTGGGGTTTGTAAACACATTACGTGCTCCCTCAGCTCCGCAAGCATCATCCC  
 TGGATGTGGATTTTACACCCCCATTCTCAAAAAACAAAGAGTATCATCAACGGGAAGTGA  
 GAGATGTTGCCCATTTAATGTGGTTCGAAAGACTCTATGTTTGGCTTCAGTGTTTTGA  
 AAAATACATCTTGTACCCAGCGCTAATTTGAATGCCCTCACTATTGATGCATTTTTTA  
 ATAAGCAATCACCGGAGACTTGGTACCCAGCTGATGATCATTGCTGGCATGAAGCTGT  
 TGCGGACATCATTCTGCAACCCGGTTTACCAGTTTATTAACCTTGAGCTTCACTGTCAT  
 CTTTTTCCACTTTGACTACAAAGATATTTACAGAGAGCTTCTTACTGGATTTCTTCATG  
 GTGTCCATTTTATTTAGCAAGGCAAGTGAATTACTTCACAAGTTACAGTTCGTCCTGA  
 CATATGTGGCTCCTTGGCAGATGGCTTGGGGTCTTTCGTTTACGTTGTTGCTCAGCT  
 CTTTGGCATTCTCGTATCCTTTCTGCCATGCTTTTCTTTTACAGACGATTGCCACATCA  
 ATCTTTTCTACCCCATTTGAGCCCATTCTTGGGAGTGTCAATTTTCATCACATCATATG  
 TCAGGCCAGTGAAATTCTGGGAGAAAACTACAGTACAAGGCGAGTGGATAATTCCAA  
 CACAAGACTGGCAGTCCAAATTGAAAGAGATCCAGGGAATGATGACAACAATCTCAAT  
 TCCATTTTTTATGAACACTTGACAAGGACCCTCCAGGAGTCCCTCTGTGGAGACTTAG  
 TTCTTGGACGTTGGGGCAACTACAGCTCTGGCGATTGCTTTATTTTGGCTTCAGATGA  
 CCTCAATGCCTTTGTTACCTGATTGAAATTGGAAATGGTCTTGTACCTTTCAACTT  
 CGAGGACTGGAATTCGAGGAACCTACTGCCAGCAGAGGGAGGTAGAAGCCATCATGG  
 AGGGCGACGAGGAGGACAGAGGCTGCTGCTGCTGCAAAACAGGCCACTTGCCCTACCT  
 GCTGTCTGCAACGCTGCCTTTACCTCCGCTGGCTCACCTGGGAAATCACGCAGACC  
 CAGTACATCCTGGAGGGCTACAGCATCCTGGACAACAACGCGGCCACCATGCTGCAGG  
 TGTTTGACCTCCGAAGGATCCTCATCCGCTACTACATCAAGAGTATAATATACTATAT  
 GGTAACGTCTCCCAAACCTCTCTCCTGGATCAAAAATGAATCACTTCTGAAGTCCCTG  
 CAGCCCTTTGCCAAGTGGCATTACATTGAGCGTGACCTTGCAATGTTCAACATTAACA  
 TTGATGATGACTACGTCCCGTGTCTCCAGGGGATCACACGAGCTAGCTTCTGCAATGT  
 TTATCTAGAATGGATTCAACACTGTGCACGGAAAAGACAAGAGCCTTCAACGACCCTG  
 GACAGTGACGAGGACTCTCCCTTGGTGACTCTGTCTTCCGCTTGTCACCTTGGGGA  
 GGAGAGCTCTGGGAACAGCCGCTCACAATATGGCCATCAGCCTGGATTCTTTCTCTGTA  
 TGGCCTCCATGTCTCTTCAAAGGTGACTTCAGAATAACAGCACGTGACGAGTGGGTA  
 TTTGCTGACATGGACCTACTGCATAAAGTTGTAGCTCCAGCTATCAGGATGTCCCTGA  
 AACTTCACCAGGACCAGTTCACTTGCCCTGACGAGTATGAAGACCCAGCAGTCCCTCTA  
 CGAGGCCATCCAGTCCTTCGAGAAGAAGGTGGTCACTGTCACGAGGGCGACCCGGCC  
 TGGCGGGGCGCAGTGCTGTCCAACAAGGAAGAGCTGCTCACCTTGGCGCACGTGGTGG  
 ACGAGGGTGCCGACGAGTACAAGGTCATCATGCTCCACAGAAGCTTCTTGAGCTTCAA  
 GGTGATCAAGGTTAACAAGAATGCGTCCGAGGACTTTGGGCGGGCAGCAGCAGGAG  
 CTTATATTTCTTCGCAACCGAATCCGGAGCGCGGCAGTATCCGAAACAATAAGCAGG  
 TCCTGCGGAACCTTGATTAACCTCTCTGCGATCAGCCCTGGGGTACCCCATGTATGT  
 CTCCCCACTAACCACATCCTACCTAGGGACACACAGGCAGCTGAAGAACATCTGGGGT  
 GGACCCATCACTTTGGACAGAATTAGGACCTGGTTCTGGACCAAGTGGGTAAGGATGC  
 GGAAGGATTGCAATGCCCGCCAGCACAGTGGCGGCAACATTGAAGACGTGGACGGAGG  
 AGGGGCCCCGACGACAGGTGGCAACAATGCCCGCAATGGTGGCAGCCAGGAGAGCAGC  
 GCAGAACAGCCCAGAAAAGGCGGTGCTCAGCACGGGGTGTCTATCCTGTGAAGGGACAC  
 AGAGAACAGGCAGGAGGAAAGGCAGGAGCCAGTCCGTGCAGGCACACTCAGCGCTAAG  
 CCAAAGGCCGCCCATGCTGAGCTCATCTGGCCCCATCTTAGAGAGCCGCCAACATTC

	CTCCAGACGTCCACCTCAGTGCACGAGCTGGCCCAGAGGCTCTCGGGCAGCCGGCTCT CCTTGACACGCCTCGGCCACGTCCCTGCACTCTCAGCCCCGCCCCGTCACCACCACCGG CCACCTGAGTGTCCGTGAGCGGGCCGAGGCGCTCATCAGGTCCAGCCTGGGCTCCTCC ACCAGCTCCACCCTGAGCTTCTCTTCGGCAAGAGGAGCTTTTCCAGCGCGCTCGTCA TTTCCGGACTCTCTGCTGCGGAGGGGGGCAATACCAGTGACACCCAGTCATCCAGCAG CGTCAACATCGTGATGGGCCCCCTCAGCCAGGGCTGCCAGCCAGGCCACTCGGGTAAGG GGCTGGGCAGGGCTCACCAGGACAGGCTGGGATGGTGGCACGGGCTCCTGGCCTGAGC GTGGCACCTGCCTTGCGTTCCACCCCTTCTGCCTGCAGAACCCCATCCCCTTCTCTAT GGGGCTCCCAGAGTGACAAAGGACAGTGATTAGACACGAAGTGGCTTAGCTGCTCTTG AAAGCAGACAAGATACAGAGCAGATATCCT		
	ORF Start: ATG at 16	ORF Stop: TGA at 4306	
	SEQ ID NO: 72	1430 aa	MW at 160787.0 Da
NOV15a, CG112869-01 Protein Sequence	MHVFIDEHGEGERSCYLKSGNQKEGPLQPLPSNNDCLSQAREMQVSSSSTTTSESQD PSSGDPVAVSALQQQLLMVARRTQSETPRHVSQDLEASSCSSTQGFCKDMWVLLF PGKWKVWYDRLTLLALLDRTEDEIKENVLAILLIVLVSLLGFLTLTSGFCKDMWVLLF CLVMASQYSLKSVQDPASPIHGNQIITYSRPIYFCVLCGLILLDTGAKARHPP SYVVYGLKLFSPVFLQSARDYLIVFLYCFPAISLLGLFPQINTFCTYLLQIDMLFFG GSAVSGITSVAVSVARSVLAALLHAVCFSAVKEPWSMQHIPALFSAFCGLLVALSYH LSRQSSDPSVLFSTFRSFIQCRLFPKFLHQNLAESAADPLPKMKDSVVRHLRLKDWL IVCAVVAVLSFAVSASTVFLSLQPFLSIVLFALAGAVGFVTHYVLPQLRKHPHWMWIS HPILKNKEYHQREVRDVAHLMWFERLYVWLQCFEKYILYPALILNALTIDAFLISNHR RLGTQLMI IAGMKLLRTSFCNPVYQFINLSFTVIFHFDDYKDISESFLDDFMVSILF SKASELLHKLQFVLTYVAPWQMAWGSSPHVFAQLFAIPRILSAMLFFQTIATISIFSTP LSPFLGSVIFITSYVRPVKFWKKNYSTRRVDNSNTRLAVQIERDPGNDDNNLNSIFYE HLTRTLQESLCGDLVLGRWGNYS SDCFILASDDLNAFVHLIEIGNGLVTFQLRGLEF RGTYCQQREVEAIMEGDEEDRGCCCCPKGHLPHLLSCNAAFHLRWLTWEITQTQYILE GYSILDNNAATMLQVFDLRRILIRYYIKSIIYYMVTSPKLLSWIKNESLLKSLQPFK WHYIERDLAMFNINIDDDYVPCLGITRASFCNVYLEWQHCAKCRQEPSTTLDSDED SPLVTL SFALCTLGRRALGTAHNMAISLDSFLYGLHVLFFKGDFRITARDEWVFADMD LLHKVVAPAIRMSLKLHQDQFTCPDEYEDPAVLYEAIQSFEKKVVICHEGDPARGAV LSNKEELLTLRHVVDEGADEYKVI MLHRSFSLFKVIKVNKECVRGLWAGQQQELIFLR NRNPERGSIQNNKQVLRNLINSSCDQPLGYPMYVSPLTTSYLGTHRQLKNIWGGPITL DRIRTFWTKWVRMRKDCNARQHS GGNIEDVDGGGAPTTGGNNAPNGGSQESSAEQPR KGGAQHGVSSCEGTQRTGRRKGRS QSVQAHSAQSQRPPMLSSSGP ILESRQTFLQTST SVHELAQRLSGSRLSLHASATSLHSQPPPVTTTGHLSVRERAEALIRSSLSGSSTSSTL SFLFGKRSFSSALVISGLSAAEGGNTSDTQSSSVNIVMGPSARAASQATRVRGWAGL TRTGWDGGTGSWPERGTCLAFPPFCLQNP IFFSMGLPE		

Further analysis of the NOV15a protein yielded the following properties shown in Table 15B.

Table 15B. Protein Sequence Properties NOV15a	
PSort analysis:	0.8000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15C.



**Table 15C. Geneseq Results for NOV15a**

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY57927	Human transmembrane protein HTMPN-51 - Homo sapiens, 777 aa. [WO9961471-A2, 02-DEC-1999]	664..1430 11..777	765/767 (99%) 765/767 (99%)	0.0
AAB01381	Neuron-associated protein - Homo sapiens, 796 aa. [WO200034477-A2, 15-JUN-2000]	529..1263 1..752	467/758 (61%) 574/758 (75%)	0.0
AAU91404	Human secreted protein sequence #57 - Homo sapiens, 595 aa. [WO200216388-A1, 28-FEB-2002]	261..840 2..581	374/588 (63%) 463/588 (78%)	0.0
AAU91356	Human secreted protein sequence #9 - Homo sapiens, 577 aa. [WO200216388-A1, 28-FEB-2002]	279..840 2..563	364/570 (63%) 451/570 (78%)	0.0
AAM79539	Human protein SEQ ID NO 3185 - Homo sapiens, 1397 aa. [WO200157190-A2, 09-AUG-2001]	89..684 80..674	333/603 (55%) 440/603 (72%)	0.0

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15D.

**Table 15D. Public BLASTP Results for NOV15a**

Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O43162	KIAA0435 protein - Homo sapiens (Human), 777 aa.	664..1430 11..777	767/767 (100%) 767/767 (100%)	0.0
Q8TEP4	FLJ00149 protein - Homo sapiens (Human), 792 aa (fragment).	664..1385 14..735	720/722 (99%) 722/722 (99%)	0.0
Q96RV3	Pecanex-like protein 1 - Homo sapiens (Human), 2341 aa.	89..1387 952..2248	738/1316 (56%) 941/1316 (71%)	0.0
Q9QYC1	Pecanex 1 - Mus musculus (Mouse), 1446 aa.	89..1371 57..1340	737/1303 (56%) 932/1303 (70%)	0.0
Q98UF7	Pecanex - Fugu rubripes (Japanese pufferfish) (Takifugu rubripes), 1703 aa.	97..1299 371..1533	722/1208 (59%) 898/1208 (73%)	0.0

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15E.

Table 15E. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

**Example 16.**

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 73	1344 bp	
NOV16a, CG113377-01 DNA Sequence	GATAAAGATGGCAATGTCTCTCATCCAAGCGTGCTGCAGTCTGGCTCTCTCAACATGG CTGCTTTCCTTTTGTTCGTGCATCTGCTCTGCCTGGACTTTACCGTGGCCGAGAAGG AGGAATGGTACACCGCCTTCGTGAACATCACCTACGCCGAGCCCGCGCCGGACCCCGG GGCCGGGGCGGCGGGCGGCGGGAGCTGCACACGGAGAAGACGGAGTGCGGG CGCTACGGAGAGCACTCGCCCCAAGCAGGACGCCCCGCGGGGAGGTGGTCATGGCCAGCT CGGCCCACGACCGCCTGGCCTGCGACCCCCAACACCAAGTTCGCCGCCCCGACCCGCGG CAAGAACTGGATAGCCCTCATCCCCAAGGGCAACTGCACGTACAGGGATAAGATCCGG AACGCGTTTCCTGCAGAACGCCTCAGCCGTGGTCATCTTCAACGTGGGCTCCAACACCA ACGAGACCATCACCATGCCCCACGCGGGTGTAGAAGACATCGTGGCCATAATGATTCC TGAGCCAAAAGGGAAGGAGATAGTAAGCCTGCTGGAAGAAACATCACCCTGACAATG TACATCACCATCGGAACCCGGAACCTGCAGAAATATGTGAGCCGCACTTCGGTTGTGT TTGTCTCCATCTCCTTCATTGTCTGATGATCATTTCCCTCGCATGGCTCGTCTTTTA TTACATCCAGAGGTTTTCGATATGCAATGCCAGGGATAGGAACCAGCGCCGACTGGGG GATGCAGCAAAGAAAGCCATCAGCAAACCTCCAGATCAGGACCATCAAGAAGGGTGACA AGGAAACAGAGTCTGATTTTGACAACCTGTGCAGTTTGTATTGAAGGGTACAAGCCCAA TGACGTTGTCCGATCCTGCCCTGCCGGCATCTTTTCCACAAGTCCTGTGTTGACCCC TGGCTTCTAGACCATCGTACCTGTCCCATGTGCAAGATGAACATTCTTAAAGCCCTAG GGATCCCGCCCAATGCCGACTGCATGGACGACTTGCCCACTGACTTCGAGGGCTCTCT GGGAGGTCCACCCACCAACCAGATCACAGGTGCCAGCGACACAACAGTGAATGAAAGT TCAGTCACTTTGGACCCTGCTGTCCGACTGTGGGAGCCTTGACAGGTGGTCCAGGATA CAGACCCCATCCCCCAGGAGGGAGACGTCATCTTTACTACTAACAGTGAGCAGGAGCC AGCTGTAAGCAGTGATTCTGACATTTCTTGATCATGGCAATGGAGGTTGGACTGTCT GATGTAGAACTTTCCACTGACCAGGACTGTGAAGAAGTGAAATCTTGAAACGACAAAT CCAGAAGCAA		
	ORF Start: ATG at 8	ORF Stop: TGA at 1322	
	SEQ ID NO: 74	438 aa	MW at 48071.3 Da
NOV16a, CG113377-01 Protein Sequence	MAMSLIQACCSLALSTWLLSFCFVHLLCLDFTVAEKEEWYAFVNITYAEPAPDPGAG AAGGGGAELHTEKTECGRYGEHSPKQDARGEVVMASAHDRACDPNTKFAAPTRGKN WIALIPKGNCTYRDKIRNAFLQNASAVVIFNVGSNTNETITMPHAGVEDIVAIMIPEP KGKEIVSLLERNITVTMYITIGTRNLQKYVSRTSVVFVSI SFIVLMIISLAWLVFYYI QRFYANARDRNQRRLGDAAKKAISKLQIRTIKKGDKETESDFDNCVCIEGYKPNV VRILPCRHLFHKSCVDPWLLDHRTPMCKMNIKALGIPPNADCMDDLPTDFEGLGG PPTNQITGASDTTVNESSVTLDPAVRTVGALQVVQDTDPI PQEGDVIFTTNSQEPV SSDSISLIMAMEVGLSDVELSTDQDCEEVKS		

Further analysis of the NOV16a protein yielded the following properties shown in

5 Table 16B.

<b>Table 16B. Protein Sequence Properties NOV16a</b>	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1080 probability located in nucleus
SignalP analysis:	Cleavage site between residues 35 and 36

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16C.

<b>Table 16C. Geneseq Results for NOV16a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV16a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAU74919	Human goliath protein sequence - Homo sapiens, 462 aa. [WO200193681-A1, 13-DEC-2001]	1..438 67..462	396/438 (90%) 396/438 (90%)	0.0
AAB41793	Human ORFX ORF1557 polypeptide sequence SEQ ID NO:3114 - Homo sapiens, 210 aa. [WO200058473-A2, 05-OCT-2000]	135..343 2..210	207/209 (99%) 207/209 (99%)	e-118
ABB90389	Human polypeptide SEQ ID NO 2765 - Homo sapiens, 419 aa. [WO200190304-A2, 29-NOV-2001]	37..401 32..385	198/368 (53%) 249/368 (66%)	e-105
AAB88558	Human hydrophobic domain containing protein clone HP03424 #2 - Homo sapiens, 419 aa. [WO200112660-A2, 22-FEB-2001]	37..401 32..385	198/368 (53%) 249/368 (66%)	e-105
AAU74921	Mouse gl protein sequence - Mus sp, 419 aa. [WO200193681-A1, 13-DEC-2001]	37..401 32..385	196/368 (53%) 247/368 (66%)	e-104

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16D.

<b>Table 16D. Public BLASTP Results for NOV16a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV16a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9ULK6	KIAA1214 protein - Homo sapiens (Human), 462 aa (fragment).	1..438 67..462	396/438 (90%) 396/438 (90%)	0.0
CAC33273	Sequence 22 from Patent WO0112660 -	37..401	198/368 (53%)	e-104

	Homo sapiens (Human), 419 aa.	32..385	249/368 (66%)	
Q8VEM1	G1-related zinc finger protein - Mus musculus (Mouse), 419 aa.	37..401 32..385	197/368 (53%) 247/368 (66%)	e-104
Q9QZQ6	G1-related zinc finger protein - Mus musculus (Mouse), 419 aa.	37..401 32..385	196/368 (53%) 247/368 (66%)	e-104
Q9P0J9	Goliath protein (Likely ortholog of mouse g1-related zinc finger protein) - Homo sapiens (Human), 276 aa.	158..401 1..242	145/244 (59%) 178/244 (72%)	3e-77

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16E.

Table 16E. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PA	81..183	26/115 (23%) 77/115 (67%)	7.1e-18
zf-C3HC4	278..318	14/54 (26%) 31/54 (57%)	1.8e-10
PHD	277..321	12/51 (24%) 29/51 (57%)	0.35

#### Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis		
	SEQ ID NO: 75	1419 bp
NOV17a, CG113730-01 DNA Sequence	GCTGCTGAGGCCAGGATATAAGGGCTGGAGGTGCTGCTTTCAGGCCTGGCCAGCCCA CCATGACAGCCCACTGCCTGCCCTTCCTTCTGCACGCCTGGTGGGCCCTACTCCAGGC GGGTGCTGCGACGGTGGCCACTGCGCTCCTGCGTACGCGGGGGCAGCCCTCGTCGCCA TCCCCCTTGCGGTACATGCTGAGCCTCTACCGCGACCCGCTGCCGAGGGCAGACATCA TCCGCAGCCTACAGGCAGAAGATGTGGCAGTGGATGGGCAGAACTGGACGTTTGCTTT TGACTTCTCCTTCCTGAGCCAACAAGAGGATCTGGCATGGGCTGAGCTCCGGCTGCAG CTGTCCAGCCCTGTGGACCTCCCCACTGAGGGCTCACTTGCCATTGAGATTTTCCACC AGCCAAAGCCCGACACAGAGCAGGCTTCAGACAGCTGCTTAGAGCGTTTCAGATGGA CCTATTCACGTGCACTTTGTCCCAGGTCACTTTTCTTGGGCAGCATGGTTTTGGAG GTGACCAGGCCTCTCTCCAAGTGGCTGAAGCACCCCTGGGGCCCTGGAGAAGCAGATGT CCAGGGTAGCTGGAGAGTGTGCGCGCGGCCCCCACACCGCCTGCCACCAATGTGCT CCTTATGCTCTACTCCAACCTCTCGCAGGAGCAGAGGCAGCTGGGTGGGTCCACCTTG CTGTGGGAAGCCGAGAGCTCCTGGCGGGCCCAGGAGGGACAGCTGTCTTGGGAGTGGG GCAAGAGGCACCGTCGACATCACTTGCCAGACAGAAGTCAACTGTGTGCGGAAGGTCAA GTTCCAGGTGGACTTCAACCTGATCGGATGGGGCTCCTGGATCATCTACCCCAAGCAG TACAACGCCTATCGCTGTGAGGGCGAGTGTCTTAATCCTGTTGGGGAGGAGTTTCATC CGACCAACCATGCATACATCCAGAGTCTGCTGAAACGTTACCAGCCCCACCGAGTCCC TTCCACTTGTGTGCCCCAGTGAAGACCAAGCCGCTGAGCATGCTGTATGTGGATAAT GGCAGAGTGCTCCTAGATCACCATAAAGACATGATCGTGAAGAATGTGGGTGCCTCT	

	GATGACATCCTGGAGGGAGACTGGATTTCCTGCACTCTGGAAGGCTGGGAAACTCCT GGAAGACATGATAACCATCTAATCCAGTAAGGAGAAACAGAGAGGGGCAAAGTTGCTC TGCCCAACCAGAACTGAAGAGGAGGGGCTGCCACTCTGTAAATGAAGGGCTCAGTGGA GTCTGGCCAAGCACAGAGGCTGCTGTGAGGAAGAGGGAGGAAGAAGCCTGTGCAGGGG GCTGGCTGGATGTTCTCTTTACTGAAAAGACAGTGGCAAGGAAAAGCACAAGTGCATG AGTTCTTTACTGGATTTTTTAAAAACC		
	ORF Start: ATG at 61	ORF Stop: TGA at 1102	
	SEQ ID NO: 76	347 aa	MW at 39560.8 Da
NOV17a, CG113730-01 Protein Sequence	MHAHCLPFLHAWWALLQAGAATVATALLRTRGQPSSPSPLAYMLSLYRDPLPRADI RSLQAEDVAVDGQNWTFADFSLFQQEDLAWAELRLQLSSPVDLPTEGSLAIEIFHQ PKPDTEQASDSCLERFQMDLFTVTLFSLGSMVLEVTPLSKWLKHPGALEKQMS RVAGECWPRPPTPPATNVLLMLYSNLSQEQRQLGGSTLLWEAESSWRAQEGQLSWEWG KRHRRHHLPLDRSQLCRKVKFQVDFNLIGWGSWIIYPKQYNAYRCEGECPNPVGEEFHP TNHAYIQSLLKRYQPHRVPSTCCAPVKTPLSMLYVDNGRVLLDHHKDMIVEECGL		
	SEQ ID NO: 77	954 bp	
NOV17b, 210982580 DNA Sequence	GGATCCCAGCCCTCGTCGCCATCCCCTCTGGCGTACATGCTGAGCCTCTACCGCGACC CGCTGCCGAGGGCAGACATCATCCGAGCCTACAGGCAGAAGATGTGGCAGTGGATGG GCAGAACTGGACGTTTGCTTTTGACTTCTCCTTCTGAGCCAACAAGAGGATCTGGCA TGGGCTGAGCTCCGGCTGCAGCTGTCCAGCCCTGTGGACCTCCCCACTGAGGGCTCAC TTGCCATTGAGATTTTCCACCAGCCAAAGCCCGACACAGAGCAGGCTTCAGACAGCTG CTTAGAGCGGTTTCAGATGGACCTATTCACTGTCACTTTGTCCCAGGTCACCTTTTCC TTGGGCAGCATGGTTTTTGAGGTGACCAGGCCTCTCTCCAAGTGGCTGAAGCACCCTG GGGCCCTGGAGAAGCAGATGTCCAGGGTAGCTGGAGAGTGTGGCCACGGCCCCCAC ACCGCCTGCCACCAATGTGCTCCTTATGCTCTACTCCAACCTCTCGCAGGAGCAGAGG CAGCTGGGTGGGTCCACCTTGCTGTGGGAAGCCGAGAGCTCCTGGCGGGCCCAGGAGG GACAGCTGTCTGGGAGTGGGGCAAGAGGCACCGTCGACATCACTTGCCAGACAGAAG TCAACTGTGTGCGGAAGGTCAAGTTCAGGTGGACTTCAACCTGATCGGATGGGGCTCC TGGATCATCTACCCCAAGCAGTACAACGCCTATCGCTGTGAGGGCAGTGTCTTAATC CTGTTGGGGAGGAGTTTCATCCGACCAACCATGCATACATCCAGAGTCTGCTGAAACG TTACCAGCCCCACCGAGTTCCTTCCACTTGTTGTGCCCCAGTGAAGACCAAGCCGCTG AGCATGCTGTATGTGGATAATGGCAGAGTGTCTCTAGATCACCATAAAGACATGATCG TGAAGAATGTGGGTGCCTCCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 78	318 aa	MW at 36367.0 Da
NOV17b, 210982580 Protein Sequence	GSQPSSPSPLAYMLSLYRDPLPRADI RSLQAEDVAVDGQNWTFADFSLFQQEDLA WAEELRLQLSSPVDLPTEGSLAIEIFHQPKPDTEQASDSCLERFQMDLFTVTLFSL LGSMVLEVTPLSKWLKHPGALEKQMSRVAGECWPRPPTPPATNVLLMLYSNLSQEQR QLGGSTLLWEAESSWRAQEGQLSWEWGKRHRRHHLPLDRSQLCRKVKFQVDFNLIGWGS WIIYPKQYNAYRCEGECPNPVGEEFHPNTHAYIQSLLKRYQPHRVPSTCCAPVKTPL SMLYVDNGRVLLDHHKDMIVEECGLLE		
	SEQ ID NO: 79	579 bp	
NOV17c, CG113794-02 DNA Sequence	ATGGTCCCCGGCGCCGCGGGCTGGTGTGTCTCGTGCTCTGGCTCCCCGCGTGCGTCG CGGCCCACGGCTTCCGTATCCATGATTATTTGTACTTTCAAGTGTGAGTCTTGGGGA CATTCGATACATCTTCACAGCCACACCTGCCAAGGACTTTGGTGGTATCTTTACACA AGGTATGAGCAGATTACCTTGTCCCCGCTGAACCTCCAGAGGCCTGCGGGGAACCTCA GCAACGGTTTCTTCATCCAGGACCAGATCGCTCTGGTGGAGAGTGGGGGCTGCTCCCT CCTCTCCAAGACTCGGGTGGTCCAGGAGCACGGCGGGCGGGCGGTGATCATCTCTGAC AATGCGGTTGACAATGACAGCTTCTATGTGGCGATGATCCAGGACAGTACCCAGCGCA CAGCTGACATCCCCGCCCTCTTCTGCTCGGCCGAGACGGCTACATGATCCGCCGCTC TCTGGAACAGCCTGGGCTGCCATGGGCCATCATTTCCATCCCAGTCAATGTCACCAGC ATCCCCACCTTTGAGCTGCAGCAACCGTCTGGTCTCTTCTGGTAGAAGGGCGATTCC		
	ORF Start: ATG at 1	ORF Stop: TAG at 565	
	SEQ ID NO: 80	188 aa	MW at 20831.6 Da

NOV17c, CG113794-02 Protein Sequence	MVPGAAGWCCLVLWLPACVAAHGFRIDYLYFQVLSPGDIRYIFTATPAKDFGGIFHT RYEQIHLVPAEPPEACGELSNGFFIQDQIALVESGGCSLLSKTRVVQEHGGRAVIISD NAVDNDSFYVAMIQDSTQRTADIPALFLLGRDGYMIRRSLEQPLPWAIIISIPVNVTS IPTFELQQPSWSFW
--------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 17B.

<b>Table 17B. Comparison of NOV17a against NOV17b and NOV17c.</b>		
<b>Protein Sequence</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV17b	34..347 3..316	314/314 (100%) 314/314 (100%)
NOV17c	340..346 89..95	4/7 (57%) 5/7 (71%)

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

<b>Table 17C. Protein Sequence Properties NOV17a</b>	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 34 and 35

5

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

<b>Table 17D. Geneseq Results for NOV17a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAY03849	Human nodal protein - Homo sapiens, 283 aa. [WO9909198-A1, 25-FEB-1999]	65..347 1..283	282/283 (99%) 282/283 (99%)	e-172
AAW56477	Amino acid sequence of human bone morphogenetic protein-16 (BMP-16) - Homo sapiens, 280 aa. [WO9812322-A1, 26-MAR-1998]	68..347 1..280	279/280 (99%) 279/280 (99%)	e-170
AAY03851	Murine nodal protein - Mus sp, 354 aa. [WO9909198-A1, 25-FEB-1999]	1..347 1..354	279/355 (78%) 298/355 (83%)	e-160
AAW84595	Amino acid sequence of the human Tango-78 protein - Homo sapiens, 169 aa.	134..297 1..164	163/164 (99%) 163/164 (99%)	2e-97

	[WO9906427-A1, 11-FEB-1999]			
AAY16702	WO9914235 Seq ID No: 155 - Unidentified, 101 aa. [WO9914235-A1, 25-MAR-1999]	247..347 1..101	99/101 (98%) 101/101 (99%)	1e-58

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

<b>Table 17E. Public BLASTP Results for NOV17a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q96S42	Nodal-related protein - Homo sapiens (Human), 347 aa.	1..347 1..347	346/347 (99%) 346/347 (99%)	0.0
P43021	Nodal precursor - Mus musculus (Mouse), 354 aa.	1..347 1..354	279/355 (78%) 298/355 (83%)	e-160
O13048	Xnr-4 - Xenopus laevis (African clawed frog), 402 aa.	31..346 72..401	123/344 (35%) 170/344 (48%)	2e-47
O13144	Nodal-related-2 (ZNR-2) - Brachydanio rerio (Zebrafish) (Zebrafish), 392 aa.	43..346 58..391	123/347 (35%) 171/347 (48%)	1e-46
P87358	ZNR-1 (CYCLOPS precursor) - Brachydanio rerio (Zebrafish) (Zebrafish), 501 aa.	243..347 397..501	71/105 (67%) 86/105 (81%)	1e-41

Pfam analysis predicts that the NOV17a protein contains the domains shown in the Table 17F.

<b>Table 17F. Domain Analysis of NOV17a</b>			
<b>Pfam Domain</b>	<b>NOV17a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
TGFb_propeptide	4..213	43/256 (17%) 122/256 (48%)	0.028
TGF-beta	244..347	46/112 (41%) 73/112 (65%)	1.5e-34

## 5 Example 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A. Note that the NOV18e nucleic acid (SEQ ID NO:121) is the reverse complement of the NOV18a residues 247-349 (SEQ ID NO:81). The NOV18e polypeptide contains additional amino acids at the ends of the ORF assembly

that are encoded by restriction endonuclease sites incorporated into amplification primers, as described in Example B.

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 81	1056 bp	
NOV18a, CG115187- 01 DNA Sequence	CACCATGCATCAGTCCCTGACTCAGCAGCGGTCCAGCGACATGTCCCTGCCCGATTCC ATGGGTGCATTCAATCGGAGGAAACGAAACTCCATCTATGTCAACCGTGACTTTGCTTA TTGTGTCCGTGTTAATTCTCACAGTGGGCCTTGCTGCAACCACCAGGACCCAGAATGT GACTGTAGGAGGTTATTACCCCGGAGTTATTCTCGGCTTTGGATCGTTCCTTGGAAATC ATTGGATCAAACCTTATTGAGAACAAAAGGCAGATGCTGGTGGCTTCTATCGTGTTTA TCAGCTTTTGGTGTGATTGCGGCTTTTTGTTGTGCCATAGTTGACGGGGTCTTTGCTGC CAGACACATTGATCTGAAACCACTCTACGCTAACCGGTGCCATTATGTTCCCAAGACA TCACAGAAGGAAGCTGAGGAGGTGATAAGTTCCTCAACCAAAAATTCTCCTTCCACGA GGGTTATGAGGAACCTTACCCAGGCAGCTAGAGAGGTAACTGCCCTCACCTCAGCCG TGAATTCTGCACACCTCGCATCCGGGGCAACACCTGCTTCTGCTGTGACCTCTACAAC TGTGGCAACCGGTGGAGATCACTGGTGGGTACTACGAATACATCGATGTCAGCAGTT GCCAAGATATCATCCACCTCTACCACCTGCTCTGGTCTGCCACCATCCTCAACATTGT TGGCCTGTTTCTTGGGCATCATCACTGCCGCTGTCTTGGAGGCTTTAAGGACATGAAC CCAACCTCTCCAGCACTGAACTGTTCTGTTGAAAATACCCATCCAACAGTTTCTTACT ATGCTCATCCCCAAGTGGCATCCTACAATACCTACTACCATAGCCCTCCTCACCTGCC ACCATATTCTGCTTATGACTTTTACGATTCCGGTGTCTTTCCATCCTCCCCCTCCCTCT GGACTTTCTGATGAGCCCCAGTCTGCCTCTCCCTCACCCAGCTACATGTGGTCTCTCAA GTGCACCGCCCCGTTACTCTCCACCCTACTATCCACCTTTTGAAAAGCCACCACCTTA CAGTCCCTAAAG		
	ORF Start: ATG at 5	ORF Stop: TAA at 1052	
	SEQ ID NO: 82	349 aa	MW at 38448.4 Da
NOV18a, CG115187- 01 Protein Sequence	MHQSLTQQRSSDMSLPDSMGAFNRRKRNSIYVTVTLIVSVLILTVGLAATTRTQNV VGGYYPGVILGFGSFLGIIGSNLIENKRQMLVASIVFISFGVIAAFCCAIVDGVFAAR HIDLKPLYANRCHYVPKTSQKEAEVVISSTKNSPSTRVMRNLQAAAREVNCPLHSRE FCTPRIRGNTCFCCDLYNCGNRVEITGGYIEYIDVSSQCDIHLHYLLWSATILNIVG LFLGIITAAVLGGFKDMNP TL PALNCSVENTHPTVSYYAHPQVASYNTYYHSPPHLPP YSAYDFQHSQVFPSSPPSGLSDEPQSASPSPSYMWSSSAPPRYSPPYPPFEKPPPYSP P		
	SEQ ID NO: 83	1049 bp	
NOV18b, CG115187- 02 DNA Sequence	CATCAGTCCCTGACTCAGCAGCGGTCCAGCGACATGTCCCTGCCCGATTCCATGGGTG CATTCAATCGGAGGAAACGAAACTCCATCTATGTCAACCGTGACTTTGCTTATTGTGTC CGTGTAAATTCTCACAGTGGGCCTTGCTGCAACCACCAGGACCCAGAATGTGACTGTA GGAGGTTATTACCCCGGAGTTATTCTCGGCTTTGGATCGTTCCTTGGAAATCATTGGAT CAAACCTTATTGAGAACAAAAGGCAGATGCTGGTGGCTTCTATCGTGTTCATCAGCTT TGGTGTGATTGCGGCTTTTTGTTGTGCCATAGTTGACGGGGTCTTTGCTGCCAGACAC ATTGATCTGAAACCACTCTACGCTAACCGGTGCCATTATGTTCCCAAGACATCACAGA AGGAAGCTGAGGAGGTGATAAGTTCTCAACCAAAAATTCTCCTTCCACGAGGGTTAT GAGGAACCTTACCCAGGCAGCTAGAGAGGTAACTGCCCTCACCTCAGCCGTGAATTC TGCACACCTCGCATCCGGGGCAACACCTGCTTCTGCTGTGACCTCTACAACGTGGCA ACCGGGTGGAGATCACTGGTGGGTACTACGAATACATCGATGTCAGCAGTTGCCAAGA TATCATCCACCTCTACCACCTGCTCTGGTCTGCCACCATCCTCAACATTGTTGGCCTG TTCTTGGGCATCATCACTGCCGCTGTCTTGGAGGCTTTAAGGACATGAACCCAACTC TCCCAGCACTGAACTGTTCTGTTGAAAATACCCATCCAACAGTTTCTTACTATGCTCA TCCCCAAGTGGCATCCTACAATACCTACTACCATAGCCCTCCTCACCTGCCACCATAT TCTGCTTATGACTTTTACGATTCCGGTGTCTTTCCATCCTCCCCCTCCCTCTGGACTTT CTGATGAGCCCCAGTCTGCCTCTCCCTCACCCAGCTACATGTGGTCTCAAGTGCACC GCCCCGTTACTCTCCACCCTACTATCCACCTTTTGAAAAGCCACCACCTTACAGTCCC TAAAG		
	ORF Start: ATG at 34	ORF Stop: TAA at 1045	



	SEQ ID NO: 84	337 aa	MW at 37048.9 Da
NOV18b, CG115187- 02 Protein Sequence	MSLPDSMGAFNRRKRNSIYVTVTLLIVSVLILTVGLAATTRTQNVTVGGYYPGVILGF GSFLGIIGSNLIENKRQMLVASIVFISFGVIAAFCCAIVDGVFAARHIDLKPLYANRC HYVPKTSQKEAEEVISSSTKNPSTRVMRNLQAAREVNCPHLSREFCTPRIRGNTCF CCDLYNCGNRVEITGGYYEYIDVSSCQDIHLYHLLWSATILNIVGLFLGIITA AVLG GFKDMNPTLPALNCSVENTHPTVSYAHQPQVASYNTYYHSPHLPYPYSAYDFQHSQGVF PSSPPSGLSDEPQSASPSPSYMWSSSAPPYSPPYPPFEKPPYPSP		
	SEQ ID NO: 85	980 bp	
NOV18c, CG115187- 03 DNA Sequence	ATGCATCAGTCCCTGACTCAGCAGCGGTCCAGCGACATGTCCCTGCCCCGATTCCATGG GAGCATTCAATCGGAGGAAACGAACTCCATCTATGTCACCGTGACTTTGCTTATTGT GTCCGTGTTAATTCTCACAGTGGGCCTTGCTGCAACCACCAGGACCCAGAATGTGACT GTAGGAGGTTATTACCCCGGAGTTATTCTCGGCTTTGGATCGTTCCTTGAATCATTG GATCAAACCTTATTGAGAACAAGGCAGATGCTGGTGGCTTCTATCGTGTTCATCAG CTTTGGTGTGATTGCGGCTTTTGTGTGCCATAGTTGACGGGTCTTTGCTGCCAGA CACATTGATCTGAAACCACTCTACGCTAACCAGGTGCCATTATGTTCCCAAGACATCAC AGAAGGAAGCTGAGGAGGTTAACTGCCCTCACCTCAGCCGTGAATTCTGCACACCTCG CATCCGGGGCAACACCTGCTTCTGCTGTGACCTCTACAACGTGGCAACCGGGTGGAG ATCACTGGTGGGTACTACGAATACATCGATGTCAGCAGTTGCCAAGATATCATCCACC TCTACCACCTGCTCTGGTCTGCCACCATCCTCAACATTGTTGGCCCCGTTTCTGGGCAT CATCACTGCCGCTGTCTTGGAGGCTTTAAGGACATGAACCCAACCTCTCCAGCACTG AACTGTTCTGTTGAAAATACCCATCCAACAGTTTCTTACTATGCTCATCCCCAAGTGG CATCCTACAATACCTACTACCATAGCCCTCCTCACCTGCCACCATATTCTGCTTATGA CTTTCAGCATTCCGGTGTCTTCCATCCTCCCCTCCTCCTGGAATTTCTGATGAGCCC CAGTCTGCCTCTCCCTCACCCAGCTACATGTGGTCTCAAGTGCACCGCCCCGTTACT CTCCACCCTACTATCCACCTTTTGAAGGCCACCACCTTACAGTCCCTAAAG		
	ORF Start: ATG at 1	ORF Stop: TAA at 976	
	SEQ ID NO: 86	325 aa	MW at 35816.4 Da
NOV18c, CG115187- 03 Protein Sequence	MHQSLTQQRSSDMSLPDSMGAFNRRKRNSIYVTVTLLIVSVLILTVGLAATTRTQNV VGGYYPGVILGFGSFLGIIGSNLIENKRQMLVASIVFISFGVIAAFCCAIVDGVFAAR HIDLKPLYANRCHYVPKTSQKEAEEVNCPHLSREFCTPRIRGNTCFCCDLYNCGNRVE ITGGYYEYIDVSSCQDIHLYHLLWSATILNIVGPFLGIITA AVLGGFKDMNPTLPAL NCSVENTHPTVSYAHQPQVASYNTYYHSPHLPYPYSAYDFQHSQGVFPSSPPSGLSDEP QSASPSPSYMWSSSAPPYSPPYPPFEKPPYPSP		
	SEQ ID NO: 87	847 bp	
NOV18d, 262770580 DNA Sequence	CACCGGATCCGCAACCACCAGGACCCAGAATGTGACTGTAGGAGGTTATTACCCCGGA GTTATTCTCGGCTTTGGATCGTTCCTTGAATCATTGGATCAAACCTTATTGAGAACA AAGGCAGATGCTGGTGGCTTCTATCGTGTTCATCAGCTTTGGTGTGATTGCGGCTTT TTGTTGTGCCATAGTTGACGGGTCTTTGCTGCCAGACACATTGATCTGAAACCACTC TACGCTAACCAGGTGCCATTATGTTCCCAAGACATCACAGAAGGAAGCTGAGGAGGTTA ACTGCCCTCACCTCAGCCGTGAATTCTGCACACCTCGCATCCGGGGCAACACCTGCTT CTGCTGTGACCTCTACAACGTGGCAACCGGTGGAGATCACTGGTGGGTACTACGAA TACATCGATGTCAGCAGTTGCCAAGATATCATCCACCTCTACCACCTGCTCTGGTCTG CCACCATCCTCAACATTGTTGGCCTGTTCTTGGGCATCATCACTGCCGCTGTCTTGG AGGCTTTAAGGACATGAACCCAACCTCTCCAGCACTGAAGTGTCTGTTGAAAATACC CATCCAACAGTTTCTTACTATGCTCATCCCCAAGTGGCATCCTACAATACCTACTACC ATAGCCCTCCTCACCTGCCACCATATCTGCTTATGACTTTCAGCATTCCGGTGTCTT TCCATCCTCCCCTCCTCTGGAATTTCTGATGAGCCCCAGTCTGCCTCTCCCTCACCC AGCTACATGTGGTCTCAAGTGCACCGCCCCGTTACTCTCCACCCTACTATCCACCTT TTGAAAAGGCCACCACCTTACAGTCCCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 88	282 aa	MW at 30945.7D
NOV18d, 262770580 Protein	TGSATTRTQNVTVGGYYPGVILGFGSFLGIIGSNLIENKRQMLVASIVFISFGVIAAF CCAIVDGVFAARHIDLKPLYANRCHYVPKTSQKEAEEVNCPHLSREFCTPRIRGNTCF CCDLYNCGNRVEITGGYYEYIDVSSCQDIHLYHLLWSATILNIVGLFLGIITA AVLG		

Sequence	GFKDMNPTLPALNCSVENTHPTVSYAHQPQVASYNTYYHSPPHLPPYSAYDFQHSGVF PSSPPSGLSDEPQSASPSPSYMWSSSAPPRYSPPYPPFEKPPPYSPLG		
	SEQ ID NO: 121	328 bp	
NOV18e, 257788219 -rev DNA Sequence; (Frame -2)	GCCCTCGAGGGGACTGTAAGGTGGTGGCTTTTCAAAGGTGGATAGTAGGGTGGAGAGTAA CGGGGCGGTGCACTTGAGGACCACATGTAGCTGGGTGAGGGAGAGGCAGACTGGGGCTCAT CAGAAAGTCCAGAGGGAGGGGAGGATGGAAAGACACCGGAATGCTGAAAGTCATAAGCAGA ATATGGTGGCAGGTGAGGAGGGCTATGGTAGTAGGTATTGTAGGATGCCACTTGGGGATGA GCATAGTAAGAACTGTTGGATGGGTATTTTCAACAGAACAGTTCAGTGCTGGGAGAGTTG GGTTCATGTCCTTGGATCCGGTG		
	ORF Start: at 328	ORF Stop: 2	
	SEQ ID NO: 122	109 aa	MW at 11964.41 Da
NOV18e, 257788219 Protein Sequence	TGSKDMNPTLPALNCSVENTHPTVSYAHQPQVASYNTYYHSPPHLPPYSAYDFQHSGVFP SSPPSGLSDEPQSASPSPSYMWSSSAPPRYSPPYPPFEKPPPYSPLG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

**Table 18B. Comparison of NOV18a against NOV18b through NOV18d.**

Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	13..315 1..303	257/303 (84%) 257/303 (84%)
NOV18c	1..315 1..291	244/315 (77%) 244/315 (77%)
NOV18d	49..315 3..245	215/267 (80%) 216/267 (80%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 50 and 51

- 5 A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq	Protein/Organism/Length [Patent #,	NOV18a	Identities/	Expect

Identifier	Date]	Residues/ Match Residues	Similarities for the Matched Region	Value
AAB31671	Amino acid sequence of a human protein having a hydrophobic domain - Homo sapiens, 166 aa. [WO200104297-A2, 18-JAN-2001]	13..130 9..126	83/118 (70%) 101/118 (85%)	1e-42
AAE03793	Human gene 13 encoded secreted protein fragment, SEQ ID NO:63 - Homo sapiens, 150 aa. [WO200132837-A1, 10-MAY-2001]	13..148 5..143	88/143 (61%) 110/143 (76%)	9e-41
AAE03776	Human gene 13 encoded secreted protein HELEN05, SEQ ID NO:46 - Homo sapiens, 71 aa. [WO200132837-A1, 10-MAY-2001]	88..148 1..64	37/68 (54%) 47/68 (68%)	6e-10
ABG06803	Novel human diagnostic protein #6794 - Homo sapiens, 106 aa. [WO200175067-A2, 11-OCT-2001]	271..349 8..78	28/79 (35%) 39/79 (48%)	4e-05
ABG06803	Novel human diagnostic protein #6794 - Homo sapiens, 106 aa. [WO200175067-A2, 11-OCT-2001]	271..349 8..78	28/79 (35%) 39/79 (48%)	4e-05

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BE63	Hypothetical 38.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 349 aa.	1..349 1..349	346/349 (99%) 347/349 (99%)	0.0
Q9NWN8	CDNA FLJ20716 fis, clone HEP19742 - Homo sapiens (Human), 185 aa.	166..349 2..185	184/184 (100%) 184/184 (100%)	e-113
Q8WV15	Hypothetical 34.6 kDa protein - Homo sapiens (Human), 326 aa.	13..349 9..326	173/343 (50%) 221/343 (63%)	3e-85
CAC28404	Sequence 24 from Patent WO0104297 - Homo sapiens (Human), 166 aa.	13..130 9..126	83/118 (70%) 101/118 (85%)	2e-42
Q9ZWT0	Extensin - Adiantum capillus-veneris (Fern), 207 aa.	264..349 46..126	31/87 (35%) 40/87 (45%)	4e-05

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

**Example 19.**

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 89	1941 bp	
NOV19a, CG115540-01 DNA Sequence	ATGGAGGGTGGCGACCCACCCCAACTCCACAGGGACAGAAGAAGCTCCTGCCTCAGG ACCGCCCTAGACACTGCCCTGTGGACCCCCTCATCTGGCTGTTTATTGTTTCTTTC TAAGCTGGTAAATGGCCCCTTGGACGGCGCGGCAAGCTTGGTAGAAGAGGCGACCCCTG GTCTTCCAGGGCAATCAGGACGAGATGGCTACCCGGGACCCCTGGGTTTGGATGGCAA GCCTGGACTTTTCAGGCCCGAAAGGGGAAAAGGGAGACCAAGGACAAGATGGAGCTGCT GGGCCTCCGGGGCCCCCTGGACCTCCTGGGGCCCCGGGGCCCTCCTGGCGACACTGGGA AAGATGGCCCCAGGGGAGCACAAGGCCAGCGGGCCCCAAAGGAGAGCCCGGACAAGA CGGCGAGATGGGCCCAAAGGGACCCCCAGGGCCCCAAGGGTGAGCCTGGAGTACCTGGA AAGAAGATGCCAGGAGCAGACTGGTGTGCTGGGAAGTCCAGAGGAGGGAGGGGCCAC TGGCCACCCGAGGGTCTGACCGGCAAGCCCCAGGTGTCTTCTCTCAGGGCGACGATG GGACACCAAGCCAGCCTGGACCACCAGGGCCCCAAGGGGGCCTCACTCTCTGCCCTGTC CCCAAGCCAGGAACCTGGGTGTCTATCCTCATGCCTTGCTCCCCCAACCCCTCGCAACAG CCACCAAATCCTGGCCAGCCAGTCTCCAAATGTCCCTTGAGCCCCCTGCGTGCCCCA AGGCGAGCCAGGGAGCATGGGGCCTCGGGGAGAGAACGGTGTGGACGGTGCCCCAGGA CCGAAGCTGCACCTCTGGCTGCAAATGCATGTCTCCACAGGGGGAGCCTGGCCACCGA GGCGCGGATGGAGCTGCAGGGCCCCGGGGTGCCCCAGGCCTCAAGGGCGAGCAGGGAG ACACAGTGGTGATCGACTATGATGGCAGGATCTTGATGCCCTCAAGGTAGTGTTCCT GGGGCCTCCCGGACCACAGGGGCCCCAGGGCCACCAGGATCCCTGGAGCCAAGGGC GAGCTTGGATTGCCCGGTGCCCCAGGAATCGATGGAGAGAAGGTCTCTGGGCCTTTCA TTTCCTTGGTGATGCCAGTGCCTGGTATTGGGCTCTGTGGCCCCAAAGGACAGAAAGG AGACCCAGGAGAGCCTGGGCCAGCAGGACTCAAAGGGGAAGCAGGCGAGATGGGCTTG TCCGGCCTCCCGGTGCTGGACACAAAGGACTCACAGGCCATTGCCGTCTTGACAGGGCG CTGACGGCCTCAAGGGGGAGAAGGGGGAGTCGGCATCTGACAGCCTACAGGAGAGCCT GGCTCAGCTCATAGTGGAGCCAGGGCCCCCTGGCCCCCTGGCCCCCAGGCCCGATG GGCCTCCAGGGAATCCAGGGTCCCAAGGGCTTGATGGAGCAAAGGGAGAGAAGGGTG CGTCGGGTGAGAGAGGCCCCAGCGGCCTGCCTGGGCCAGTTTGGCCCAACCGGGCCTTAT TGGGCTGCCAGGAACCAAGGAGAGAAGGGCAGACCCGGGGAGCCAGGACTAGATGGT TTCCCTGGACCCCGAGGAGAGAAAGGTGATCGGAGCGAGCGTGGAGAGAAGGGAGAAC GAGGGGTCCCCGGCCGGAAGGAGTGAAGGGCCAGAAGGGCGAGCCGGGACCACCAGG CCTGGACCAGCCGTGTCCCGTGGGCCCCGACGGGCTGCCTGTGCCTGGCTGCTGGCAT AAGAACCTGCTCCCGCAAACTCTGGAGTCCCTGGGACACACCCTATCCAAGAAGACC CAGGGGTGGAACAGCGGCTGCTGTTGCTCCTGGCCTCATCAGCCTCCAACTCAACCA CAACCAGCTGCCTCTGCAGTTGGACAAGACTTGGCCCCCGGACAAGACTCGCCCAGCA CTTGCGGCTGGGCCCGGGGAGCAGTGA		
	ORF Start: ATG at 1	ORF Stop: TGA at 1939	
	SEQ ID NO: 90	646 aa	MW at 66246.7 Da
NOV19a, CG115540-01 Protein Sequence	MEGGDPTPTPQGQKLLPQDRPRHCPVDPLIWLFIILSKLVNGLDGAASLVEEATL VLQGNQDEMATRDPWVWMAASLDFQARKGKRETKDKMELLGLRGPLDLLGPGALLATLG KMAPGEHKAQRAPKESPKTKARWAQRDPQGPRVSLEYLERRCQEQTGVLGSP EEGGAH WPPEGLTGKPVQVSSPQGDDGTSPSQPGPPGPKGASLSALSPSQELGVILMPCSPNPSQQ		

	PPNPGQPVSKMSLEPLRCPKASQGAWGLGERTVWTVPQDRSCTSGCKCMSPQGEPPGHR GADGAAGPRGAPGLKGEQGDVVIDYDGRILDALKVVFLGPPGPQGGPPGPIPGAKG ELGLPGAPGIDGEKVSGPFISLVMPVPGIGLCGPKGQKGDGPGEPPAGLKGEGEMGL SGLPVLDTKDSQAI AVLQADGLKGEKGESASDSLQESLAQLIVEPGPPGPPGPPGPM GLQGIQGPGLDGAKEKGASGERGPSGLPGVPVGPGLIGLPGTKGEKGRPGEPGLDG FPGPRGEKGDRSERGEKGERGVPRKGVKQKGEPPGLDQPCPVGPDGLFVPGCWH KNLLPQNSGVPGTHPIQEDPGVEQRLLLLLASSASKLNHNQLPLQLDKTWPPDKTRPA LAAGPGEQ
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Further analysis of the NOV19a protein yielded the following properties shown in Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 45 and 46

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB43239	Human ORFX ORF3003 polypeptide sequence SEQ ID NO:6006 - Homo sapiens, 200 aa. [WO200058473-A2, 05-OCT-2000]	349..581 1..200	189/233 (81%) 189/233 (81%)	e-106
AAG63332	Amino acid sequence of human collagen-like protein CLAC - Homo sapiens, 654 aa. [WO200158943-A1, 16-AUG-2001]	98..581 234..654	208/495 (42%) 247/495 (49%)	7e-83
AAG63343	Amino acid sequence of murine collagen-like protein CLAC - Mus sp, 666 aa. [WO200158943-A1, 16-AUG-2001]	98..581 234..666	205/509 (40%) 240/509 (46%)	1e-82
AAR53257	Human collagen (Type V) - Homo sapiens, 1838 aa. [JP06105687-A, 19-APR-1994]	98..576 1135..1538	176/486 (36%) 209/486 (42%)	2e-61
AAY08305	Human collagen IX alpha-2 chain protein - Homo sapiens, 705 aa. [WO9921011-A1, 29-APR-1999]	98..605 30..518	188/545 (34%) 233/545 (42%)	4e-60

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NT93	Hypothetical 19.5 kDa protein - Homo sapiens (Human), 201 aa (fragment).	349..581 1..201	201/233 (86%) 201/233 (86%)	e-115
Q99MQ5	Collagen-like alzheimer amyloid plaque component precursor type I - Mus musculus (Mouse), 666 aa.	98..581 234..666	205/509 (40%) 240/509 (46%)	3e-82
Q9NQ52	Type XIII collagen - Homo sapiens (Human), 717 aa.	159..581 263..717	198/488 (40%) 235/488 (47%)	3e-75
O70575	Collagen type XIII alpha-1 chain - Mus musculus (Mouse), 739 aa.	159..581 270..739	197/495 (39%) 233/495 (46%)	1e-74
Q14035	Alpha-1 type XIII collagen - Homo sapiens (Human), 623 aa.	159..581 170..623	192/488 (39%) 231/488 (46%)	3e-70

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen	283..341	23/60 (38%) 41/60 (68%)	0.0033
Collagen	342..401	22/60 (37%) 36/60 (60%)	0.0014
Collagen	448..506	32/60 (53%) 43/60 (72%)	1.4e-07
Collagen	507..566	27/60 (45%) 46/60 (77%)	1.1e-10

#### Example 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis		
	SEQ ID NO: 91	1247 bp
NOV20a, CG118689-01 DNA Sequence	GCCCTACCGTGTGCGCAGAAAGAGGAGGCGCTTGCCTTCAGCTTGTGGGAAATCCCGA AGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCCCTGCTGATTTTGGAAA TGTGATTATTGGTTGTTGCGGCATTGCCCTGACTGCGGAGTGCATCTTCTTTGTATCT GACCAACACAGCCTCTACCCACTGCTTGAAGCCACCGACAACGATGACATCTATGGGG	

	CTGCCTGGATCGGCATATTTGTGGGCATCTGCCTCTTCTGCCTGTCTGTTCTAGGCAT TGTAGGCATCATGAAGTCCAGCAGGAAAATTCTTCTGGCGTATTTTCATTCTGATGTTT ATAGTATATGCCTTTGAAGTGGCATCTTGTATCACAGCAGCAACACAACGAGACTTTT TCACACCCAACCTCTTCTGAAGCAGATGCTAGAGAGGTACCAAAACAACAGCCCTCC AAACAATGATGACCAGTGGAAAAACAATGGAGTCACCAAAACCTGGGACAGGCTCATG CTCCAGGACAATTGCTGTGGCGTAAATGGTCCATCAGACTGGCAAAAATACACATCTG CCTTCCGGACTGAGAATAATGATGCTGACTATCCCTGGCCTCGTCAATGCTGTGTTAT GAACAATCTTAAAGAACCTCTCAACCTGGAGGCTTGTAAGTAGGCGTGCCTGGTTTT TATCACAATCAGTTTTGGGTTCTCTGGGTACCATGTTCTACTGGAGCAGAATTGAAT ATTAAGCATAAAGTGTGGCCACCATAACCTCCTTCCCCGAGTGACTCTGGATTGGTGC TGGAAACCAGCTCTCTCCTAATATTCCACGTTTGTGCCCCACACTAACGTGTGTGTT ACATTGCCAAGTCAGATGGTACGGACTTCTTTAGGATCTCAGGCTTCTGCAGTTCTC ATGACTCCTACTTTTTCATCCTAGTCTAGCATTCTGCAACATTTATATAGACTGTTGAA AGGAGAATTTGAAAAATGCATAATAACTACTTCCATCCCTGCTTATTTTTAATTTGGG AAAATAAATACATTGCAAGGAAAAACAAAAAAAAGGGCGGCCCCCGATTATTGAGGGG TCCCGAGCCCGAACTCGTAACCATGTAAAACCCGTCCCCGGGGTAAAAATTGTAATCC CCCCACAATTCCCCAAAACATAGGGCCCCGAAGCCTAAAGTTTAAAACCTGGGGGGG CCTAAGGAGTTTACCCAAACTCCCTTCT		
	ORF Start: ATG at 61	ORF Stop: TAA at 757	
	SEQ ID NO: 92	232 aa	MW at 26502.3 Da
NOV20a, CG118689-01 Protein Sequence	MAKDNSTVRCFQGLLIFGNVIIGCCGIALTAECIFFVSDQHSPLYLLEATDNDDIYGA AWIGIFVGICLFLSVLGIVGIMKSSRKILLAYFILMFIVYAFEVASCITAATQRDF TPNLFKQMLERYQNNSPPNDDQWKNGVTKTWDRMLQDNCCGVNGPSDWQKYTSA FRTENNDADYPWPRQCCVMNLEPLNLEACKLGVPGFYHNQFVLLGTMFYWSRIEY		
	SEQ ID NO: 93	851 bp	
NOV20b, CG118689-02 DNA Sequence	GAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCCTGCTGATTTTTGGA AATGTGATTATTGGTTGTTGCGGCATTGCCCTGACTGCGGAGTGCATCTTCTTTGTAT CTGACCAACACAGCCTCTACCCACTGCTTGAAGCCACCGACAACGATGACATCTATGG GGCTGCCTGGATCGGCATATTTGTGGGCATCTGCCTCTTCTGCCTGTCTGTTCTAGGC ATTGTAGGCATCATGAAGTCCAGCAGGAAAATTCTTCTGGCGTATTTTCATTCTGATGT TTATAGTATATGCCTTTGAAGTGGCATCTTGTATCACAGCAGCAACACAACGAGACTT TATGCTAGAGAGGTACCAAAACAACAGCCCTCCAAATAATGATGACCAGTGGAAAAAC AATGGAGTCACCAAAACCTGGGACAGGCTCATGCTCCAGGACAATTGCTGTGGCGTAA ATGGTCCATCAGACTGGCAAAAATACACATCTGCCTTCCGGACTGAGAATAATGATGC TGACTATCCCTGGCCTCGTCAATGCTGTGTTATGAACAATCTTAAAGAACCTCTCAAC CTGGAGGCTTGTAAGTAGGCGTGCCTGGTTTTTATCACAATCAGGGCTGCTATGAAC TGATCTCTGGTCCAATGAACCGACACGCCTGGGGGGTTCCTGGTTTGGATTGTCAT TCTCTGCTGGACTTTTTGGGTTCTCCTGGGTACCATGTTCTACTGGAGCAGAATTGAA TATTAGGCATAAAGTGTGGCCACCATAACCTCCTTCCCCGAGTGACTCTGGATTGGT GCTGGAACCAGCTCTCTCCTAATATTCACGTTTGTGCC		
	ORF Start: ATG at 5	ORF Stop: TAG at 758	
	SEQ ID NO: 94	251 aa	MW at 28581.7 Da
NOV20b, CG118689-02 Protein Sequence	MAKDNSTVRCFQGLLIFGNVIIGCCGIALTAECIFFVSDQHSPLYLLEATDNDDIYGA AWIGIFVGICLFLSVLGIVGIMKSSRKILLAYFILMFIVYAFEVASCITAATQRDFM LERYQNNSPPNDDQWKNGVTKTWDRMLQDNCCGVNGPSDWQKYTSFRTENNDAD YPWPRQCCVMNLEPLNLEACKLGVPGFYHNQGCYELISGPMNRHAWGVAVFGFAIL CWTFWLLGTMFYWSRIEY		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b.		
Protein Sequence	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV20b	1..232	223/260 (85%)
	1..251	223/260 (85%)

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 31 and 32

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY94419	Human TM4P-1 protein - Homo sapiens, 260 aa. [WO200026243-A2, 11-MAY-2000]	1..232 1..260	232/260 (89%) 232/260 (89%)	e-137
AAE10871	Bovine uroplakin Ib protein - Bos sp, 260 aa. [US6290959-B1, 18-SEP-2001]	1..232 1..260	214/260 (82%) 225/260 (86%)	e-126
AAE10870	Bovine uroplakin Ia protein - Bos sp, 258 aa. [US6290959-B1, 18-SEP-2001]	13..208 18..208	81/198 (40%) 116/198 (57%)	1e-42
AAM48320	Human tetraspan - Homo sapiens, 248 aa. [FR2809734-A1, 07-DEC-2001]	4..223 2..214	67/229 (29%) 111/229 (48%)	2e-16
AAB49503	Clone HCE1K90 #1 - Homo sapiens, 248 aa. [WO200070076-A1, 23-NOV-2000]	4..223 2..214	67/229 (29%) 111/229 (48%)	2e-16

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value



O75841	Uroplakin Ib (UPIb) - Homo sapiens (Human), 259 aa.	2..232 1..259	231/259 (89%) 231/259 (89%)	e-136
A41531	TGFbeta-regulated protein TI-1 - American mink, 260 aa.	1..232 1..260	217/260 (83%) 228/260 (87%)	e-129
P30413	Uroplakin Ib (UPIb) (TI 1 protein) - Mustela vison (American mink), 259 aa.	2..232 1..259	216/259 (83%) 227/259 (87%)	e-128
I46081	uroplakin Ib - bovine, 260 aa.	1..232 1..260	214/260 (82%) 225/260 (86%)	e-126
P38573	Uroplakin Ib (UPIb) - Bos taurus (Bovine), 259 aa.	2..232 1..259	213/259 (82%) 224/259 (86%)	e-125

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
transmembrane4	12..225	53/256 (21%) 163/256 (64%)	2.3e-43

### Example 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis		
	SEQ ID NO: 95	1518 bp
NOV21a, CG120748-01 DNA Sequence	CGGGCATGAAGGAGGATGGAAGGGCAGGACGAGGTGTCGGCGCGGGAGCAGCACTTCC ACAGCCAAGTGCGGGAGTCCACGATATGTTTCCTTCTTTTGGCATTCTCTACGTTGT TTCCTACTTCATCATCACAAGATACAAGAGAAAATCAGATGAACAAGATGAAGAT GCCATCGTCAACAGGATTTGCTTGTGTTTGTAGCACGTTCACTCTCGCAGTGTGCTG GGGCTGTTTTGCTTTTACCCTTCTCAATCATCAGCAATGAAATCCTGCTTTCTTTTCC TCAGAACTACTATATTAGTGCTAAATGGCTCCCTGATTCATGGTTTGTGGAATCTT GCTTCCCTTTTTTCCAACCTTTGTTTATTTGTATTGATGCCCTTTGCCCTTTTCTTTC TGGAATCAGAAGGCTTTGCTGGCCTGAAAAAGGAATCCGAGCCCGCATTTTAGAGAC TTTGGTCATGCTTCTTCTTCTTGCCTTACTCATTCTTGGGATAGTGTGGGTAGCTTCA GCACTCATTGACAACGATGCCGCAAGCATGGAATCTTTATATGATCTCTGGGAGTTCT ATCTACCCTATTTATATTCTGTATATCATTGATGGGATGTTTGTACTTCTCTTGTG TACACCAGTTGGCCTTTCTCGTATGTTACAGTGATGGGTGCTAGTGAAGCCA ACAATTCTGAAGACCTGGATGAACAAATTTATATCATTACCTTAGAGGAAGAAGCAC TCCAGAGACGACTAAATGGTCTGTCTTCATCGGTGGAATACAACATAATGGAGTTGGA ACAAGAAGTTGAAATGTAAAGACTCTTAAGACAAAATAGATAGGCGAAAAAGGCT TCAGCATGGGAAAGAAATTTGGTGTATCCCGCTGTTATGGTTCCTCTTCTTATTGAGA CATCCATCTCGGTCTCTTGGTGGCTTGAATATTCTTTGCCCTATTGGTTGATGAAAC AGCAATGCCAAAAGGAACAAGGGGGCTGGAATAGGAAATGCCTCTCTTTCTACGTTT GGTTTTGTGGGAGCTGCGCTTGAAATCATTTTGATTTTCTATCTTATGGTGTCTCTG TTGTCGGCTTCTATAGCCTTCGATTTTTTGGAACTTTACTCCCAAGAAAGATGACAC AACTATGACAAAGATCATTGGAATTTGTGTGTCCATCTTGGTTTTGAGCTCTGCTCTG	

	CCTGTGATGTCGAGAACACTGGGAATCACTAGATTTGATCTACTTGGCGACTTTGGAA GGTTTAATTGGCTGGGAAATTTCTATATTGTATTATCCTACAATTTGCTTTTTGCTAT TGTGACAACATTGTGTCTGGTCCGAAAATTCACCTCTGCAGTTCGAGAAGAACTTTTC AAGGCCCTAGGTCTTCATAAACTTCACTTACCAAATACTTCAAGGGATTGAGAAACAG CCAAGCCTTCTGTAAATGGGCATCAGAAAGCACTGTGAGACGCACAGACGGCGTCTTC TGCCACCAAG		
	ORF Start: ATG at 16	ORF Stop: TGA at 1486	
	SEQ ID NO: 96	490 aa	MW at 55083.0 Da
NOV21a, CG120748-01 Protein Sequence	MEGQDEVSAREQHFHSQVRESTICFLLFAILYVVSFYIITRYKRKSDEQEDEDIVNR ISLFLSTFTLAVSAGAVLLLPFSIISNEILLSFPQNYIQLWNGSLIHGLWNLASLFS NLCLFVLMPPFAFFFLSEGFAGLKKGIRARILETLVMLLLLALLILGIVVVASALIDN DAASMESLYDLWEFYLPYLYSCISLMGCLLLLCTPVGLSRMFTVMGQLLVKPTILED LDEQIYIITLEEEALQRRNLGLSSSVEYNIMELEQELENVKTLLKTLDRRKASAWER NLVYPVMVLLLIETSIISVLLVACNILCLLVDETAMPKGTRGPGIGNASLSTFGFVGA ALEIILIFYLMVSSVVGFYSLRFFGNFTPKKDDTTMTKIIGNCVSILVLSSALPVMR TLGITRFDLLGDFGRFNWLGNFYIVLSYNLLFAIVTTLCILVRKFTSAVREELFKALGL HKLHLPNTSRDSETAKPSVNGHQKAL		

Further analysis of the NOV21a protein yielded the following properties shown in Table 21B.

Table 21B. Protein Sequence Properties NOV21a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 36 and 37

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21C.

5

Table 21C. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY91600	Human secreted protein sequence encoded by gene 9 SEQ ID NO:273 - Homo sapiens, 407 aa. [WO200006698-A1, 10-FEB-2000]	84..490 1..407	405/407 (99%) 406/407 (99%)	0.0
ABB11389	Human secreted protein homologue, SEQ ID NO:1759 - Homo sapiens, 415 aa. [WO200157188-A2, 09-AUG-2001]	85..490 9..415	393/407 (96%) 397/407 (96%)	0.0
ABB90410	Human polypeptide SEQ ID NO 2786 - Homo sapiens, 367 aa. [WO200190304-A2, 29-NOV-2001]	124..490 1..367	366/367 (99%) 367/367 (99%)	0.0
AAG75542	Human colon cancer antigen protein SEQ	174..490	315/317 (99%)	e-178

	ID NO:6306 - Homo sapiens, 345 aa. [WO200122920-A2, 05-APR-2001]	29..345	316/317 (99%)	
AAY91459	Human secreted protein sequence encoded by gene 9 SEQ ID NO:132 - Homo sapiens, 313 aa. [WO200006698- A1, 10-FEB-2000]	179..490 1..312	310/312 (99%) 311/312 (99%)	e-175

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21D.

Table 21D. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8WVP7	Hypothetical 55.1 kDa protein - Homo sapiens (Human), 490 aa.	1..490 1..490	489/490 (99%) 490/490 (99%)	0.0
Q96QZ5	Differentiation-related DIF14 long form - Homo sapiens (Human), 490 aa.	1..490 1..490	486/490 (99%) 487/490 (99%)	0.0
Q9JIT0	LMBR1 long form - Mus musculus (Mouse), 490 aa.	1..490 1..490	470/490 (95%) 483/490 (97%)	0.0
Q91WC6	Similar to limb region 1 - Mus musculus (Mouse), 463 aa.	1..490 1..463	443/490 (90%) 456/490 (92%)	0.0
Q969J4	Lipocalin-1 interacting membrane receptor (Lipocalin-interacting protein) - Homo sapiens (Human), 487 aa.	7..470 9..473	287/465 (61%) 363/465 (77%)	e-165

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21E.

Table 21E. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

## 5 Example 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 97	1217 bp
NOV22a, CG121519-01	ATGGGGGTGAGGGTAGAGTGGGGGTACGGTGAAGGGAGGCTGAGTCTAATGAAAATAT TGATGGAATACTGCTTTTGT TTTTGT TTTCCCTAGTGGGAGAGATCTCTGAGCTTTG	

DNA Sequence	TCCGGAAATCACTGATTTTTTGTGCCGGGACAAGAAGTGCATTGCATCCCACCTTCTT TGTGACTATAAGCCAGACTGCTCTGATAGGTCTGATGAAGCTCACTGTGCACATTATA CAAGCACAAAGGAAGCTGCAATTTTGAAACAAGTTCAGGAACTGGACCACAGCCTG CAGTCTTACTCAAGACTCTGAGGATGACTTGGACTGGGCCATTGGCAGCAGAATTCCT GCCAAAGCATTAAATCCAGACTCTGATCACACGCCAGGTAGTGGTCAGCACTTTCTGT ACGTCAACTCATCTGGCTCCAAGGAAGGATCCGTTGCCAGAATTACTACTTCCAAATC CTTCCCAGCAAGCCTTGGAAATGTGTACTGTTTCGGTTCTGGTTCTACATGATTGATCCC AGGAGTATGGGAATATTAAAGGTGTATACCATTGAAGAATCGGGGCTAAACATCCTCG TGTGGTCAGTGATTGGAAATAAAAGAACGGGATGGACATATGGCTCTGTGCCCTCTCTC CAGTAACAGTCCGTTTAAGGTGGCATTTGAAGCTGATTTGGATGGAAATGAGGACATC TTTATTGCTCTTGATGACATCTCTTTTACCCAGAGTGTGTGACTGGAGGTCTGTCC CAGTGCAGCCATCACCTGTGAAGCTGATCAGTTTTCTTGTATCTACACACTCCAATG TGTCCCTCTCTCAGGGAAATGTGATGGACATGAAGACTGCATAGATGGACCTGATGAA ATGGATTGTCCTCTCAGCCCCACCCCTCCACTCTGTAGTAACATGGAGTTCCCGTGTCT CTACAGACGAGTGTATACCTTCCCTCCTGCTATGCGATGGAGTGCCCGACTGCCACTT TAATGAAGATGAGCTCATCTGCTCCAACAAAAGCTGTTCTAATGGAGCTCTGGTGTGT GCCTCCTCCAACAGCTGTATCCCAGCCCACCAGCGCTGTGATGGTTTTGCCGACTGCA TGGATTTCCAGCTTGATGAGTCCAGCTGCTCCGGTACCCCATTTCCATTGAGATATTC TTGTGATATGAACCAGCAACTTAACCTGCAACACAATGAAAATATTAAAACTTGAAG		
	ORF Start: ATG at 1	ORF Stop: TGA at 1213	
	SEQ ID NO: 98	404 aa	MW at 44271.1 Da
NOV22a, CG121519-01 Protein Sequence	MGVRVEWGYGEGRLSLMKILMEYCFVSVSLVGEISELCPEITDFLCRDKKCIASHLL CDYKPDSDRSDEAHCAHYTSTTGSCNFETSSGNWTTACSLTQDSEDDLWDWAIGSRIP AKALIPDSHTPGSGQHFLYVNSSGSGKEGSVARITTSKSFASLMCTVRFWFYMDP RSMGILKVYTIIESGLNVLVSVIGNKRTGWYGSVPLSSNSPFKVAFEADLDGNEDI FIALDDISFTPECVTGGPVPVQSPCEADQFSCIYTLQCVPLSGKCDGHEDCIDGPDE MDCPLSPTPPLCSNMEFPCSTDECIPSLLLCDGVPDCHFNEDELICSNKSCSNGALVC ASSNSCIPAHQRCDGFADCMDFQLDESSCSGTPFPFRYSCDMNQQLNLQHNENIKT		

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM92795	Human digestive system antigen SEQ ID NO: 2144 - Homo sapiens, 94 aa. [WO200155314-A2, 02-AUG-2001]	113..199 7..93	86/87 (98%) 86/87 (98%)	6e-44

AAW02212	Human VLDL receptor - Homo sapiens, 873 aa. [WO9626286-A1, 29-AUG-1996]	258..388 33..161	54/135 (40%) 69/135 (51%)	5e-20
AAR74691	Human very low density lipoprotein receptor - Homo sapiens, 846 aa. [WO9513374-A2, 18-MAY-1995]	258..388 6..134	54/135 (40%) 69/135 (51%)	5e-20
ABG23265	Novel human diagnostic protein #23256 - Homo sapiens, 4436 aa. [WO200175067-A2, 11-OCT-2001]	250..377 303..429	55/132 (41%) 63/132 (47%)	7e-20
AAB31889	Amino acid sequence of a human protein - Homo sapiens, 4393 aa. [WO200105422-A2, 25-JAN-2001]	250..377 277..403	55/132 (41%) 63/132 (47%)	7e-20

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P20063	Low-density lipoprotein receptor (LDL receptor) - Oryctolagus cuniculus (Rabbit), 837 aa (fragment).	258..383 55..177	55/129 (42%) 68/129 (52%)	1e-20
Q9J118	Low density lipoprotein receptor related protein LRP1B/LRP-DIT - Mus musculus (Mouse), 4599 aa.	244..377 3539..3667	51/134 (38%) 68/134 (50%)	5e-20
JE0237	apolipoprotein E receptor 2 precursor - mouse, 996 aa.	248..388 33..168	55/144 (38%) 71/144 (49%)	1e-19
Q924X6	Apolipoprotein E receptor 2 precursor - Mus musculus (Mouse), 996 aa.	248..388 33..168	55/144 (38%) 71/144 (49%)	1e-19
P98155	Very low-density lipoprotein receptor precursor (VLDL receptor) - Homo sapiens (Human), 873 aa.	258..388 33..161	54/135 (40%) 69/135 (51%)	1e-19

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ldl_recept_a	37..76	16/43 (37%) 27/43 (63%)	7.6e-08

MAM	84..247	58/179 (32%) 115/179 (64%)	2.9e-23
Idl_recept_a	256..295	18/43 (42%) 29/43 (67%)	4.9e-07
Idl_recept_a	300..338	16/43 (37%) 29/43 (67%)	4.2e-09
Idl_recept_a	339..379	15/44 (34%) 29/44 (66%)	0.0025

**Example 23.**

The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis			
	SEQ ID NO: 99	654 bp	
NOV23a, CG122176-01 DNA Sequence	<b>ATGAGGACTGAGAAGGCTGTACCCCCACAGTCACAACCTCCTCCGCCTGTGGCTGGGCT</b> GCGTCTGCTTCGCGCTGGTGCAGGCGGACAGTCCCTCAGCCCCAGTGAACGTACACGTCAGGTCACCTCAAGGCCAACTCTGCAGTGGTGAGCTGGGATGTTCTGGAGGATGAGGTTGTCATCGGATTTGCCATCTCCAGCAGAAGAAGGATGTGCGGATGCTGCGCTTCATCCAGGAGGTGAACACCACCACCCGCTCATGTGCCCTCTGGGACCTGGAGGAGGATACGGAGTACATAGTCCACGTGCAGGCCATCTCCATTTCAGGGCCAGAGCCCAGCCAGCGAGCCTGTGCTCTTCAAGACCCCGCGTGAGGCTGAGAAGATGGCCTCCAAGAACAAGATGAGGTAACCATGAAAGAGATGGGGAGGAACCAACAGCTGCGGACAGGCGAGGTGCTGATCATCGTCGTGGTCCCTGTTTCATGTGGGCAGGTGTCATTGCCCTCTTCTGCCGCCAGTATGACATCATCAAGGACAATGAACCCAATAACAACAAGGAAAAAACCAGAGTGCATCAGAAA CCAGCACACCAGAGCACCAGGGCGGGGGGCTTCTCCGCAGCAAGGTGTTCCAAACAAGCCCTCAGTGAACATCA		
	ORF Start: ATG at 1	ORF Stop: TGA at 646	
	SEQ ID NO: 100	215 aa	MW at 24129.5 Da
NOV23a, CG122176-01 Protein Sequence	<b>MRTEKAVPPQSQLLRLWLGCVCFALVQADSPSAPVNVTVRHLKANSAVVSWDVLEDEV</b> VIGFAISQKKDVRMLRFIQEVNTTTRSCALWDLEEDTEYIVHVQAISIQGQSPASEPVLFKTPREAEMASKNKDEVTMKEMGRNQQLRTGEVLIIVVLFMWAGVIALFCRQYDI IKDNEPNNNKEKTKSASETSTPEHQGGGLLRSKVFQTS PQ		

Further analysis of the NOV23a protein yielded the following properties shown in

## 5 Table 23B.

Table 23B. Protein Sequence Properties NOV23a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 29 and 30

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23C.

<b>Table 23C. Geneseq Results for NOV23a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV23a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABG08028	Novel human diagnostic protein #8019 - Homo sapiens, 161 aa. [WO200175067-A2, 11-OCT-2001]	31..108 83..160	46/78 (58%) 62/78 (78%)	2e-20
ABG08028	Novel human diagnostic protein #8019 - Homo sapiens, 161 aa. [WO200175067-A2, 11-OCT-2001]	31..108 83..160	46/78 (58%) 62/78 (78%)	2e-20
AAE05252	Mouse Nope (neighbour of punc ell) extracellular domain - Mus musculus, 932 aa. [WO200149714-A2, 12-JUL-2001]	25..153 401..538	44/138 (31%) 69/138 (49%)	3e-08
AAE05251	Mouse Nope (neighbour of punc ell) protein - Mus musculus, 1252 aa. [WO200149714-A2, 12-JUL-2001]	25..153 422..559	44/138 (31%) 69/138 (49%)	3e-08
ABG08027	Novel human diagnostic protein #8018 - Homo sapiens, 103 aa. [WO200175067-A2, 11-OCT-2001]	31..67 66..102	26/37 (70%) 31/37 (83%)	6e-08

In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23D.

<b>Table 23D. Public BLASTP Results for NOV23a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV23a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
CAC82194	Peroxisomal protein (PeP) - Mus musculus (Mouse), 209 aa.	6..209 7..209	198/204 (97%) 201/204 (98%)	e-109
Q9DB97	1500001L03Rik protein - Mus musculus (Mouse), 209 aa.	6..209 7..209	192/204 (94%) 198/204 (96%)	e-106
Q9H6D8	CDNA: FLJ22362 fis, clone HRC06544 - Homo sapiens (Human), 234 aa.	26..194 42..213	92/172 (53%) 128/172 (73%)	1e-45
CAC51121	Sequence 3 from Patent WO0149714 - Mus musculus (Mouse), 932 aa (fragment).	25..153 401..538	44/138 (31%) 69/138 (49%)	6e-08
Q9JLI1	Neighbor of Punc ell protein - Mus musculus (Mouse), 1252 aa.	25..153 422..559	44/138 (31%) 69/138 (49%)	6e-08

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23E.

Table 23E. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
fn3	31..114	29/87 (33%) 64/87 (74%)	1.9e-12

**Example 24.**

The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 101	3930 bp
NOV24a, CG122691-01 DNA Sequence	CCCGAGCACCATGAGCTCCGGAGACCCTGCACACCTCGGCCTCTGCCTCTGGCTGTGG CTGGGCGCCACCCTGGGAAGAGAGCAAGTTCAAGCAAGCGGTCTCCTGAGGCTGGCTG TGCTGCCTGAGGACCGGCTGCAGATGAAGTGGAGAGAGTCCGAGGGGAGCGGCCTCGG CTACCTGGTGCAGGTGAAGCCCATGGCAGGGGACTCGGAACAGGAGGTGATACCTGACC ACCAAGACCCCTAAGGCCACAGTGGGGGGCCTGAGCCCCCTCCAAGGGCTACACCTTGC AGATCTTCGAGCTCACTGGCTCTGGGCGCTTCTCTGCTAGCTCGGAGGGAGTTTGTGGT TGAGGATCTGAAGAGTAGCTCCCTGGACAGGAGCAGCCAGAGGCCCTCGGCTCTGGA GCCCCGGAGCCACCCCTCCACACGGGGAGCCCAGACCCTGAGCAGGCTTCTGAGC CCCAAGTTGCCTTCACACCAAGCCAGGATCCGCGCACTCCTGGTGGGTGAGAGTGGAG AGAGACCGGCCCCCAGTTCCGCTGCCTGCCCCCGTGCTGCTGACATGGTCTTCTCTG GTGGACGGGTCTGGAGCATTTGGCCACAGTCACTTCCAGCAGGTCAAGGACTTCTCTGG CCAGTGTTCATCGCACCCCTTTGAAATCGGGCCGGATAAGGTCCAAGTAGGCTGACTCA GTACAGCGGGGATGCTCAGACTGAGTGGGACCTGAACTCCCTCAGCACCAAGGAACAG GTGCTGGCAGCTGTGCGCCGCTCCGCTACAAGGGGGGAACACGTTTACAGGCCTTG CCCTGACCCACGTGCTGGGGCAGAACCTGCAGCCGGCGGCTGGCTCCGTCCAGAGGC AGCCAAGGTGGTGATTTCTGGTGACGGACGGCAAGTCCAGGACGATGTGCACACTGCT GCCCCGTGCTCAAGGACCTGGGCGTGAACGCTTCTGCTGTGGGTGTGAAGAACGCCG ATGAGGCTGAGCTGAGGCTCCTGGCGTCCCCGCGGAGGGACATACCGTCCACAGCGT GCTGGACTTCTTGCAGCTCGGCGCGCTGGCTGGCCTGCTCAGCCGTCTCATCTGCCAG AGGCTCCAGGGTGGGAGCCCGCGGACGGGCCCAGCAGCGGCTCCAGCCCTGGACACCC TCCCTGCCCCCACCAGCCTGGTCTTGAGCCAGGTGACCTCCTCCAGCATCCGCTGTCT CTGGACTCCAGCCCCCGGCACCCCTCAAGTATCTGATCGTTTGGCAGACCTCTAGA GGTGGCACCCCCAGGGAGGTGGTGGTGGAGGGGCCCCGCGCCTCCACGGAGCTGCACA ACCTGGCCTCCCGCACAGAGTACCTGGTCTCCGTGTTCCCCATCTATGAGGGCGGGGT TGGCGAAGGCCTGCGGGGCTGGTGACCACAGCACCTCTGCCTCCGCCCCGGGCGCTG ACCCTGGCCGAGTGACGCCAGAACCGTCCACCTCACCTGGCAGCCCTCGGCCGGGG CCACCCACTACCTGGTGCATGTTCTCCTGCTTCCCCCAAGGGTGAAGAGGAGGAGCG AGAGGTGCAGGTGGGGCGGCCCCGAGGTGCTGCTGGATGGCCTGGAACCTGGCAGGGAC TATGAGGTCTCGGTGCAGAGCCTGCGAGGCCCTGAGGGCAGCGAGGCCCGGGGCATCC GTGCCAGGACCCCCACCCTGGCCCCCCCCGAGACACCTGGGCTTCTCAGACGTGAGCCA CGACGCGGCACGAGTGTCTGGGAGGGTGCCCCGAGGCTGTGCGCCTGGTCAGGGTC ACCTATGTGTCCAGCGAGGGTGGACACTCGGGGCAGACAGAGGCTCCTGGGAACGCCA CCTCGGCCATGCTGGGGCCTCTCTTCTTCTCCACCACCTACACTGTCCGTGTACCTG CCTCTACCCCTGGGGGTGGCTCCTCTACGCTGACTGGCCGGGTGACCACCAAGAAAGCT CCCAGCCCAAGCCAGCTGTCCATGACGGAGCTGCCAGGGGATGCAGTCCAGCTGGCGT GGGTGGCCGAGCCCCGTCTGGCGTGCTTGTCTACCAGATCACGTGGACGCCCTTGGG AGAGGGGAAGGCTCACGAGATCTCTGTCCAGGGAAACCTCGGCACGGCCGTCTGCCT GGCTTAGGGAGGCACACAGAGTACGACGTACCATCTTGGCCTACTACAGGACGGGG CCCGCAGTGACCTGTGTCCCTCCGCTATACCCCTCCACGGTGAGCAGGAGCCACC CTCCAACCTGGCCCTGGCCTCGAGACCCCCGACAGCTGCAGTCAAGTCAAGCGCCC CCGCTTGGCCGCGTGCTCCATTACTGGCTCACCTACGCCCCCGCCTCTGGCTTGGGAC	



	CCGAGAAATCCGTCTCTGTGCCAGGAGCCAGGAGCCACGTGACACTGCCCGACCTGCA GGCAGCCACGAAGTACAGGGTCTTGGTCTCAGCTATCTATGCAGCAGGCAGGAGTGAG GCTGTGTCTGCCACGGGCCAGACAGCCTGCCAGCCCTCCGCCCTGACGGCTCCCTCC CAGGGTTTGACCTGATGGTGGCCTTCAGCCTGGTGGAAAAGGCTTATGCGTCCATCCG GGGCGTGGCCATGGAGCCCTCTGCCTTCGGTGGGACCCCCGACCTTCACGCTCTTCAAG GACGCCCAGCTGACAAGACGGGTCACTGACGTCTACCCAGCCCCCTACCTCCAGAGC ACACCATCGTCTTCTTGTGCGCCTACTTCCCGAGACACCCCGTGAGGCCTTCGCGCT GTGGCAGATGACAGCCGAGGACTTCCAGCCCCCTCTTGGGGTCTTGCTGGATGCCGGG AAGAAGTCCCTGACCTACTTCCACCGTGACCCAGGGCTGCCTTGAGGAGGCCACCT TCGACCCGCAGGAAGTGAGGAAGATTTTCTTCGGGAGCTTCCACAAGGTGCACGTGGC TGTGGGCCGCTCCAAGGTCAAGGCTCTATGTGGACTGCCGGAAGGTGGCTGAGCGGCC CTTGGGGAGATGGGCAGCCACCCGCTGCCGGCTTCGTACGCTGGGGAGGCTGGCCA AGGCCAGGGGCCCCCGAGCAGTTTCGGCCCGTTTCAGCTCCAGATGCTGCAGATCGT GTGCAGTGACACCTGGGCCGATGAGGACCGGTGCTGTGAGCTCCCTGCCTCGAGGGAT GGAGAGACCTGCCCCGCCTTCGTGTCTGCCTGTTCTGTTCCTCAGAGACCCCTGGGC CCCCAGGACCTCAAGGACCCCCAGGCCTCCCTGGGAGGAATGGACCCAGGAGAGCA GGGCTTCCAGGGGCCAGGGGTCCACAGGGGTCAAAGGAGAGAGAAGGGAGACCATGGG CTTCAGGCTTGCAGGGCCACCCGGCCACAGGGCATCCCCGGGAGAGTTGGCCTCC AGGGACCAAAGGAATGAGAGGCCTGGAGGGAAGTCTGGCCTGCCTGGACCCCTGG CCCCAGGGGTTCCAGGGCATGGCAGGGGCCAGGGGCACTAGTGAGAGCGAGGACCT CCAGGGACCGTGGGGCCACAGGACTGCCAGGGCCAAAGGGGAACGAGGAGAGAAGG GCGAGCCGAGTCCCTTGCCACCCTCTACCAGCTTGTGAGCCAGGCCTGTGAGTCTGC CATTCAGACACACGTGTCAAAGTTCGACTCCTTCCACGAGAACACCAGGCCCCCATG CCCATCTTGAGCAGAAGCTGGAGCCGGGCACTGAGCCCTGGGGTCCCTGGCACCC GCAGCAAGGCCCTGGTTCTTGAGAATGGGGGCGTGGTGGCCGCCACCTTGAGGGCAG AGGGGAGCCTGGAGCTGTTGGTCAGATGGGCAGCCCTGGGCAGCAGGGGGCTAGCACC CAGGGCCTCTGGGAGTGACAGGACATTTTCTGCACTGCCCCGAG		
	ORF Start: ATG at 11	ORF Stop: TGA at 3902	
	SEQ ID NO: 102	1297 aa	MW at 137322.2 Da
NOV24a, CG122691-01 Protein Sequence	MSSGDP A H L G L C L W L G A T L G R E Q V Q A S G L L R L A V L P E D R L Q M K W R E S E G S G L G Y L V Q V K P M A G D S E Q E V I L T T K P K A T V G G L S P S K G Y T L Q I F E L T G S G R F L L A R R E F V V E D L K S S L D R S S Q R P L G S G A P E P T P S H T G S P D P E Q A S E P Q V A F T P S Q D P R T P G G S E W R E T G P Q F R C L P P V P A D M V F L V D G S W S I G H S H F Q Q V K D F L A S V I A P F E I G P D K V Q V G L T Q Y S G D A Q T E W D L N S L S T K E Q V L A A V R R L R Y K G G N T F T G L A L T H V L G Q N L Q P A A G L R P E A A K V V I L V T D G K S Q D D V H T A A R V L K D L G V N V F A V G V K N A D E A E L R L L A S P P R D I T V H S V L D F L Q L G A L A G L L S R L I C Q R L Q G G S P R Q G P A A P A L D T L P A P T S L V L S Q V T S S I R L S W T P A P R H P L K Y L I V W R A S R G G T P R E V V V E G P A A S T E L H N L A S R T E Y L V S V F P I Y E G G V G E G L R G L V T T A P L P P P R A L T L A A V T P R T V H L T W Q P S A G A T H Y L V R C S P A S P K G E E E E R E V Q V G R P E V L L D G L E P G R D Y E V S V Q S L R G P E G S E A R G I R A T P T L A P P R H L G F S D V S H D A A R V F W E G A P R P V R L V R V T Y V S S E G G H S G Q T E A P G N A T S A M L G P L S S T T Y T V R V T C L Y P G G G S T L T G R V T T K A P S P S Q L S M T E L P G D A V Q L A W A A A P S G V L V Y Q I T W T P L G E G K A H E I S V P G N L G T A V L P G L G R H T E Y D V T I L A Y Y R D G A R S D P V S L R Y T P S T V S R S P P S N L A L A S E T P D S L Q V S W T P P L G R V L H Y W L T Y A P A S G L G P E K S V S V P G A R S H V T L P D L Q A A T K Y R V L V S A I Y A A G R S E A V S A T G Q T A C P A L R P D G S L P G F D L M V A F S L V E K A Y A S I R G V A M E P S A F G G T P T F T L F K D A Q L T R R V S D V Y P A P L P P E H T I V F L V R L L P E T P R E A F A L W Q M T A E D F Q P L L G V L L D A G K K S L T Y F H R D P R A A L Q E A T F D P Q E V R K I F F G S F H K V H V A V G R S K V R L Y V D C R K V A E R P L G E M G S P P A A G F V T L G R L A K A R G P R S S A A F Q L Q M L Q I V C S D T W A E D R C C E L P A S R D G E T C P A F V S A C S C S S E T P G P P G P Q G P P G L P G R N G T P G E Q G F P G P R G P P G V K E K G D H L P G L Q G H P G H Q G I P G R V G L Q G P K G M R G L E G T A G L P G P P G P R G F Q G M A G A R G T S G E R G P P G T V G P T G L P G P K G E R G E K G E P Q S L A T Y Q L V S Q A C E S A I Q T H V S K F D S F H E N T R P P M P I L E Q K L E P G T E P L G S P G T R S K A L V P G E W G R G R H L E G R G E P G A V G Q M G S P G Q Q G A S T Q G L W E		

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

<b>Table 24B. Protein Sequence Properties NOV24a</b>	
PSort analysis:	0.4500 probability located in cytoplasm; 0.4409 probability located in microbody (peroxisome); 0.2395 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24C.

<b>Table 24C. Geneseq Results for NOV24a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV24a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB90762	Human Tumour Endothelial Marker polypeptide SEQ ID NO 257 - Homo sapiens, 3063 aa. [WO200210217-A2, 07-FEB-2002]	24..836 994..1841	287/872 (32%) 432/872 (48%)	e-120
AAU27790	Human full-length polypeptide sequence #115 - Homo sapiens, 3118 aa. [WO200164834-A2, 07-SEP-2001]	24..836 994..1841	287/872 (32%) 432/872 (48%)	e-120
AAU84267	Human endometrial cancer related protein, COL14A1 - Homo sapiens, 755 aa. [WO200209573-A2, 07-FEB-2002]	830..1289 184..668	207/493 (41%) 282/493 (56%)	e-112
AAB27229	Human EXMAD-7 SEQ ID NO: 7 - Homo sapiens, 795 aa. [WO200068380-A2, 16-NOV-2000]	830..1289 208..692	207/493 (41%) 282/493 (56%)	e-112
AAG73916	Human colon cancer antigen protein SEQ ID NO:4680 - Homo sapiens, 561 aa. [WO200122920-A2, 05-APR-2001]	848..1289 17..479	192/477 (40%) 269/477 (56%)	7e-99

- In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

<b>Table 24D. Public BLASTP Results for NOV24a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV24a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9BQU7	BA261N11.4.1 (KIAA1510, isoform 1) - Homo sapiens (Human), 1207 aa.	1..1201 1..1201	1164/1201 (96%) 1167/1201 (96%)	0.0
Q9P218	KIAA1510 protein - Homo sapiens	158..1297	1140/1140 (100%)	0.0

	(Human), 1140 aa (fragment).	1..1140	1140/1140 (100%)	
Q90ZA0	Collagen type XX alpha 1 precursor - Gallus gallus (Chicken), 1472 aa.	22..1295 20..1409	710/1399 (50%) 926/1399 (65%)	0.0
AAH30415	Hypothetical 80.8 kDa protein - Mus musculus (Mouse), 765 aa (fragment).	568..1292 4..760	552/759 (72%) 607/759 (79%)	0.0
Q923P2	BM401L17.1.1 (Novel protein similar to KIAA1510 (Isoform 1)) - Mus musculus (Mouse), 573 aa (fragment).	747..1292 3..568	410/579 (70%) 450/579 (76%)	0.0

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
fn3	26..108	19/87 (22%) 47/87 (54%)	0.17
vwa	186..358	84/201 (42%) 141/201 (70%)	2.4e-59
fn3	384..467	30/87 (34%) 60/87 (69%)	2.4e-12
fn3	474..552	28/86 (33%) 64/86 (74%)	4.1e-16
fn3	564..646	25/86 (29%) 54/86 (63%)	7.3e-06
fn3	654..734	23/84 (27%) 57/84 (68%)	1.5e-10
fn3	747..827	29/86 (34%) 61/86 (71%)	2.5e-14
TSPN	849..1044	50/222 (23%) 153/222 (69%)	2.1e-35
Collagen	1077..1135	34/60 (57%) 43/60 (72%)	1.4e-09
Collagen	1136..1195	29/60 (48%) 45/60 (75%)	6.9e-12

#### Example 25.

The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis			
	SEQ ID NO: 103	610 bp	
NOV25a, CG122863-01 DNA Sequence	CATGAAACAGCAGCAGTGGTGTGGGATGACTGCCAAAATGGGCACCGTGTTGTCAGGG GTCTTCACCATCATGGCCGTAGACATGTATCTCATCTTTGAACAGAAGCACCTAGGGA ATGGCAGTTGCACTGAGATCACACCAAAGTACAGGGGTGCAAGTAACATCATAAATAA CTTCATCATCTGCTGGAGTTTTAAAATCGTCCTCTTCCTGTCTTTTCATCACCATCCTC ATCAGCTGCTTCCTCCTGTACTCAGTGTATGCCCAGATCTTCAGGGGCCTGGTCATCT ACATTGTCTGGATTTTTTTTCTATGAAACTGCAAACGTCGTAATACAAATCCTCACCAA CAATGACTTTGACATTAAAGAGGTCAGAATCATGCGCTGGTTTGGCTTGGTGTCTCGT ACAGTCATGCACGTGTTTCTGGATGTTCTTTGTTCATCAACTATGCCACATAACCTACA AAAACCGGAGCCAGGGCAATATAATTTCTACAAGAGACGAATTTCTACAGCGGAGAT TCTCCACAGCAGAAATAAAAGATTATCAATTTTCGAGTGGGTTTCAGTGGCTCACACCTG GAATCCCAGTACTTTGAGAGGCAGAGGTAG		
	ORF Start: ATG at 2	ORF Stop: TAG at 608	
	SEQ ID NO: 104	202 aa	MW at 23735.7 Da
NOV25a, CG122863-01 Protein Sequence	MKQQQWCGMTAKMGTVLSGVFTIMAVDMYLIFEQKHLGNGSCTEITPKYRGASNI INN FIICWSFKIVLFLSFITILISCFLLYSVYAQIFRGLVIYIVWIFFYETANVVIQILTN NDFDIKEVRIMRWFLVSRVTVMHCFWMFFVINYAHITYKNRSQGNIIISYKRRISTAEI LHSRNKRLSISSGFSGSHLESQYFERQR		
	SEQ ID NO: 105	610 bp	
NOV25b, CG122863-02 DNA Sequence	CATGAAACAGCAGCAGTGGTGTGGGATGACTGCCAAAATGGGCACCGTGTTGTCAGGG GTCTTCACCATCATGGCCGTAGACATGTATCTCATCTTTGAACAGAAGCACCTAGGGA ATGGCAGTTGCACTGAGATCACACCAAAGTACAGGGGTGCAAGTAACATCATAAATAA CTTCATCATCTGCTGGAGTTTTAAAATCGTCCTCTTCCTGTCTTTTCATCACCATCCTC ATCAGCTGCTTCCTCCTGTACTCAGTGTATGCCCAGATCTTCAGGGGCCTGGTCATCT ACATTGTCTGGATTTTTTTTCTATGAAACTGCAAACGTCGTAATACAAATCCTCACCAA CAATGACTTTGACATTAAAGAGGTCAGAATCATGCGCTGGTTTGGCTTGGTGTCTCGT ACAGTCATGCACGTGTTTCTGGATGTTCTTTGTTCATCAACTATGCCACATAACCTACA AAAACCGGAGCCAGGGCAATATAATTTCTACAAGAGACGAATTTCTACAGCGGAGAT TCTCCACAGCAGAAATAAAAGATTATCAATTTTCGAGTGGGTTTCAGTGGCTCACACCTG GAATCCCAGTACTTTGAGAGGCAGAGGTAG		
	ORF Start: ATG at 2	ORF Stop: TAG at 608	
	SEQ ID NO: 106	202 aa	MW at 23735.7 Da
NOV25b, CG122863-02 Protein Sequence	MKQQQWCGMTAKMGTVLSGVFTIMAVDMYLIFEQKHLGNGSCTEITPKYRGASNI INN FIICWSFKIVLFLSFITILISCFLLYSVYAQIFRGLVIYIVWIFFYETANVVIQILTN NDFDIKEVRIMRWFLVSRVTVMHCFWMFFVINYAHITYKNRSQGNIIISYKRRISTAEI LHSRNKRLSISSGFSGSHLESQYFERQR		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 25B.

Table 25B. Comparison of NOV25a against NOV25b.		
Protein Sequence	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV25b	1..202	202/202 (100%)
	1..202	202/202 (100%)

Further analysis of the NOV25a protein yielded the following properties shown in Table 25C.

<b>Table 25C. Protein Sequence Properties NOV25a</b>	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 26 and 27

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25D.

<b>Table 25D. Geneseq Results for NOV25a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE10587	Human macrophage-expressed protein #12 - Homo sapiens, 127 aa. [WO200164839-A2, 07-SEP-2001]	1..123 1..123	123/123 (100%) 123/123 (100%)	2e-67
AAU08333	Hamster Beta 2 adrenergic receptor - Ceratotherium simum, 309 aa. [US6277591-B1, 21-AUG-2001]	84..156 199..268	18/73 (24%) 34/73 (45%)	0.37
AAP90550	Hamster beta-2 -adrenergic receptor - Cricetus, 390 aa. [WO8918149-A, 08-SEP-1989]	84..156 199..277	18/79 (22%) 37/79 (46%)	0.84
AAE14409	Beta-2 adrenergic receptor derived 4-transmembrane helix receptor - Unidentified, 255 aa. [WO200187976-A2, 22-NOV-2001]	116..169 186..234	15/54 (27%) 29/54 (52%)	5.6
AAG81843	S. epidermidis open reading frame protein sequence SEQ ID NO:780 - Staphylococcus epidermidis, 432 aa. [WO200134809-A2, 17-MAY-2001]	56..104 15..71	18/57 (31%) 29/57 (50%)	5.6

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25E.

<b>Table 25E. Public BLASTP Results for NOV25a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q8TC54	Similar to RIKEN cDNA 4933413N12 gene - Homo sapiens (Human), 211 aa.	1..202 1..202	202/202 (100%) 202/202 (100%)	e-116

Q9D446	4933413N12Rik protein - Mus musculus (Mouse), 189 aa.	1..170 1..167	99/170 (58%) 130/170 (76%)	7e-56
AAH29657	Similar to hypothetical gene supported by BC026012 - Homo sapiens (Human), 170 aa.	9..168 1..159	43/162 (26%) 77/162 (46%)	3e-12
T18438	hypothetical protein C0415c - malaria parasite (Plasmodium falciparum), 1532 aa.	108..159 937..1000	20/64 (31%) 37/64 (57%)	0.69
O77332	Hypothetical 204.0 kDa protein - Plasmodium falciparum (isolate 3D7), 1673 aa.	108..159 1078..1141	20/64 (31%) 37/64 (57%)	0.69

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25F.

Table 25F. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Known Matches Found			

#### Example 26.

The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis			
	SEQ ID NO: 107	617 bp	
NOV26a, CG50880-04 DNA Sequence	GTACCTGTTCTCTGCCCTGGAAGTGCCTCGTGGTCGTGTCTCTCAGGCTGCTGTTCTCT GTACCCACAGGAGTGCCCGTGCGCAGCGGAGATGCCACCTTCCCCAAAGCTATGGACA ACGTGACGGTCCGGCAGGGGGAGAGCGCCACCCTCAGGTGCACCTATTGACAACCGGGT CACCCGGGTGGCCTGGCTAAACCGCAGCACCATCCTCTATGCTGGGAATGACAAGTGG TGCCTGGATCCTCGCGTGGTCCTTCTGAGCAACACCCAAACGCAGTACAGCATCGAGA TCCAGAACGTGGATGTGTATGACGAGGGCCCTTACACCTGCTCGGTGCAGACAGACAA CCACCCAAAGACCTCTAGGGTCCACCTCATTGTGCAAGTATCTCCCAAATTTGTAGAG ATTTCTTCAGATATCTCCATTAATGAAGGGAACAATATTAGCCTCACCTGCATAGCAA CTGGTAGACCAGAGCCTACGGTTACTTGGAGACACATCTCTCCCAAAGGTCCAGGCGC CGTCAGCGAGGTGAGCAACGGCACGTGAGGAGGGCAGGCTGCGTCTGGCTGCTGCCT CTTCTGGTCGTGCACCTGCTTCTCAAATTTGATGTG		
	ORF Start: at 2	ORF Stop: TGA at 611	
	SEQ ID NO: 108	203 aa	MW at 22526.8 Da
NOV26a, CG50880-04 Protein Sequence	YLFLPWKCLVVVSLRLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRV TRVAWLNRSTILYAGNDKWCLDPRVLLSNTQTQYSIEIQNVDVYDEGPYTCVQTDN HPKTSRVHLIVQVSPKIVEISSDISINEGNNISLTCIATGRPEPTVTRHISPKGPGA VSEVSNGTSTRAGCVWLLPLLVLVHLLK		

Further analysis of the NOV26a protein yielded the following properties shown in Table 26B.

<b>Table 26B. Protein Sequence Properties NOV26a</b>	
PSort analysis:	0.9190 probability located in plasma membrane; 0.3000 probability located in lysosome (membrane); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 26 and 27

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26C.

<b>Table 26C. Geneseq Results for NOV26a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV26a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAM38713	Human polypeptide SEQ ID NO 1858 - Homo sapiens, 344 aa. [WO200153312-A1, 26-JUL-2001]	1..177 6..182	174/177 (98%) 174/177 (98%)	8e-99
ABB84844	Human PRO337 protein sequence SEQ ID NO:56 - Homo sapiens, 344 aa. [WO200200690-A2, 03-JAN-2002]	23..177 28..182	152/155 (98%) 152/155 (98%)	3e-85
AAU83654	Human PRO protein, Seq ID No 126 - Homo sapiens, 344 aa. [WO200208288-A2, 31-JAN-2002]	23..177 28..182	152/155 (98%) 152/155 (98%)	3e-85
AAB31204	Amino acid sequence of human polypeptide PRO337 - Homo sapiens, 344 aa. [WO200077037-A2, 21-DEC-2000]	23..177 28..182	152/155 (98%) 152/155 (98%)	3e-85
AAU12359	Human PRO337 polypeptide sequence - Homo sapiens, 344 aa. [WO200140466-A2, 07-JUN-2001]	23..177 28..182	152/155 (98%) 152/155 (98%)	3e-85

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26D.

<b>Table 26D. Public BLASTP Results for NOV26a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV26a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9P121	Neurotrimin - Homo sapiens (Human), 344 aa.	1..177 6..182	174/177 (98%) 174/177 (98%)	2e-98
Q62718	Neurotrimin precursor (GP65) -	1..177	174/177 (98%)	2e-98

	Rattus norvegicus (Rat), 344 aa.	6..182	174/177 (98%)	
Q99PJ0	Neurotrimin - Mus musculus (Mouse), 344 aa.	1..177 6..182	171/177 (96%) 171/177 (96%)	2e-96
Q9DGI5	CEPU-Se alpha 1 isoform - Gallus gallus (Chicken), 315 aa.	2..177 7..182	154/176 (87%) 167/176 (94%)	2e-89
O93242	CEPU-1 - Gallus gallus (Chicken), 344 aa.	2..177 7..182	154/176 (87%) 167/176 (94%)	2e-89

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26E.

Table 26E. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig	45..112	18/72 (25%) 45/72 (62%)	6.3e-08
ig	145..166	10/24 (42%) 18/24 (75%)	0.2

#### Example 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
	SEQ ID NO: 109	2005 bp
NOV27a, CG51812-03 DNA Sequence	CTATAAAAGCTGTCGGTCTCTTAAGGCTGCCAGCGCCTTGCCAAAATGGAGCTTGTA GAAGGCTCATGCCATTGACCTCTTAATTCTCTCTGTTGGCGGAGCTGACAATGGC GGAGGCTGAAGGTAAGGCAAGCTGCACAGTCAGTCTAGGGGGTGCCAATATGGCAGAG ACCCACAAAGCCATGATCCTGCAACTCAATCCCAGTGAGAAGTGCACCTGGACAATAG AAAGACCAGAAAACAAAAGCATCAGAATTATCTTTTCCTATGTCCAGAGGCTTGATCC AGATGGAAGCTGTGAAAGTGAAAACATTAAAGTCTTTGACGGAACCTCCAGCAATGGG CCTCTGCTAGGGCAAGTCTGCAGTAAAAACGACTATGTTCTGTATTTGAATCATCAT CCAGTACATTGACGTTTCAAATAGTTACTGACTCAGCAAGAATTCAAAGAAGTGTCTT TGTCTTCTACTACTTCTTCTCTTCCATTTTCAGCTATTCCAAACTGTGGCGGTTACCTG GATACCTTGGAAGGATCCTTCACCAGCCCCAATTACCCAAAGCCGCATCCTGAGCTGG CTTATTGTGTGTGGCACATACAAGTGGAGAAAGATTACAAGAAAATAGAATTGAATTG GTTTGAAACTCTTTTTCACCAGAGGCCCTCATTTCTAGCCTAGAAAATAGACAAACAG TGCAAATTTGATTTTCTTGCCATCTATGATGGCCCCCTCCACCAACTCTGGCCTGATTG GACAAGTCTGTGGCCGTGTGACTCCACCTTCGAATCGTCATCAAACCTCTGACTGT CGTGTGTCTACAGATTATGCCAATTCTTACCGGGGATTTTCTGCTTCTACACCTCA ATTTATGCAGAAAACATCAACACTACAGCATCTTTAACTTGCTCTTCTGACAGGATGA GAGTTATTATAAGCAAATCCTACCTAGAGGCTTTTAACTCTAATGGGAATAACTTGCA ACTAAAAGACCCAACTTGACAGACCAAAATATCAAATGTTGTGGAATTTTCTGTCCCT CTTAATGGATGTGGTACAATCAGAAAGGTAGTAGAAGATCAGTCAATTACTTACACCA ATATAATCACCTTTTCTGCATCCTCAACTTCTGAAGTGATCACCCGTGAGAAACAAC CCAGATTATTGTGAAGTGTGAAATGGGACATAATTCTACAGTGGAGATAATATACATA ACAGAAGATGATGTAATACAAAGTCAAAATGCACTGGGCAAATATAACACCAGCATGG CTCTTTTGAATCCAATTCATTTGAAAAGACTATACTTGAATCACCATATTATGTGGA	



	TTTGAACCAAACCTCTTTTGTTCAGTTAGTCTGCACACCTCAGATCCAAATTTGGTG GTGTTTCTTGATACCTGTAGAGCCTCTCCACCTCTGACTTTGCATCTCCAACCTACG ACCTAATCAAGAGTGGGTGTAGTCGAGATGAAACTTGTAAGGTGTATCCCTTATTTGG ACACTATGGGAGATTCCAGTTTAATGCCTTTAAATTCTTGAGAAGTATGAGCTCTGTG TATCTGCAGTGTAAAGTTTGTATATGTGATAGCAGTGACCACCAGTCTCGCTGCAATC AAGGTTGTGTCTCCAGAAGCAAACGAGACATTTCTTCATATAAATGGAAAACAGATTC CATCATAGGACCCATTCTGTCTGAAAAGGGATCGAAGTGCAAGTGGCAATTCAGGATTT CAGCATGAAACACATGCGGAAGAAACTCCAAACCAGCCTTTCAACAGTGTGCATCTGT TTTCCTTCATGGTTCTAGCTCTGAATGTGGTGACTGTAGCGACAATCACAGTGAGGCA TTTTGTAAATCAACGGGCAGACTACAAATACCAGAAGCTGCAGAACTATTA <del>ACTAACA</del> GGTCCAACCCTAAGTGAGACATGTTTCTCCAGGATGCCAAAGGAAATGCTACCTCGTG GCTACACATATTATGAATAAATGAGGAAGGGCC		
	ORF Start: ATG at 46	ORF Stop: TAA at 1906	
	SEQ ID NO: 110	620 aa	MW at 69654.9 Da
NOV27a, CG51812-03 Protein Sequence	MELVRRRLMPLTLLILSCLAELTMAEAEKGASCTVSLGGANMAETHKAMILQLNPSENC TWTIERPENKSIRIIFSIVQRLDPDGSCESENIKVFDTSSNGPLLQVCSKNDYVPV FESSSTLTQIVTDSARIQRTVFVFYFFSSISAIPNCGGYLDTLEGSFTSPNYPKP HPELAYCVWHIQVEKDYKKIELNWFETLFHQRPSPSSLEIDKQCKFDLAIYDGPSTN SGLIGQVCGRVPTPFESSNSLTVVLDSTDYANSYRGFSASYTSIYAENINTTASLTCS SDRMVRIISKSYLEAFNSNGNNLQLKDPTCRPKLSNVVEFSVPLNGCGTIRKVVEDQS ITYTNIITFSASSTSEVITRQQLQIIVKCEMGHNSTVEIIYITEDDVIQSQNALGKY NTSMALFESNSFEKTILESPIYVDLNQTLFVQVSLHTSDPNLVVFLDTCRASPTSDFA SPTYDLIKSGCSRDETKVYPLFGHYGRFQFNAFKFLRSMSSVYLQCKVLICDSSDHQ SRCNQGCVSRSKRDISSYKWKTDIIGPIRLKRDRSASGNSGFQHETHAETPNQPFN SVHLFSFMVLALNVTVATITVRHFVNQRADYKYQKLQNY		

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.2800 probability located in endoplasmic reticulum (membrane); 0.2000 probability located in lysosome (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 25 and 26

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27C.

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB80245	Human PRO257 protein - Homo sapiens, 607 aa. [WO200104311-A1, 18-JAN-2001]	1..620 1..607	598/620 (96%) 602/620 (96%)	0.0
AAU12343	Human PRO257 polypeptide sequence - Homo sapiens, 607 aa. [WO2001040466-	1..620 1..607	598/620 (96%) 602/620 (96%)	0.0

	A2, 07-JUN-2001]			
AAY13377	Amino acid sequence of protein PRO257 - Homo sapiens, 607 aa. [WO9914328-A2, 25-MAR-1999]	1..620 1..607	598/620 (96%) 602/620 (96%)	0.0
AAY25323	Human pancreatic PA153 consensus protein - Homo sapiens, 607 aa. [WO9931274-A2, 24-JUN-1999]	1..620 1..607	598/620 (96%) 602/620 (96%)	0.0
AAB07456	Protein encoded by a novel gene associated with insulin synthesis - Homo sapiens, 585 aa. [WO200040722-A2, 13-JUL-2000]	23..620 1..585	576/598 (96%) 580/598 (96%)	0.0

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O35360	Uterus-ovary specific putative transmembrane protein - Rattus norvegicus (Rat), 607 aa.	1..620 1..607	444/620 (71%) 520/620 (83%)	0.0
Q9QZT0	Estrogen-regulated protein - Rattus norvegicus (Rat), 607 aa.	1..620 1..607	443/620 (71%) 519/620 (83%)	0.0
P70412	Integral membrane-associated protein 1 - Mus musculus (Mouse), 606 aa.	1..620 1..606	429/620 (69%) 517/620 (83%)	0.0
Q9HAR7	Uterus-ovary specific putative transmembrane protein UO - Homo sapiens (Human), 357 aa.	268..620 7..357	349/353 (98%) 349/353 (98%)	0.0
Q62827	Ebnerin - Rattus norvegicus (Rat), 1290 aa.	155..602 838..1274	165/456 (36%) 270/456 (59%)	9e-79

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

Table 27E. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value
CUB	38..144	25/121 (21%) 70/121 (58%)	0.00056
CUB	155..273	39/128 (30%) 91/128 (71%)	2.1e-28

zona_pellucida	288..532	77/287 (27%) 184/287 (64%)	3.1e-38
----------------	----------	-------------------------------	---------

**Example 28.**

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis		
	SEQ ID NO: 111	14536 bp
NOV28a, CG51923-01 DNA Sequence	GGAGTTTTCACCATGACTATTGCCCTGCTGGGTTTGGCCATATTCTTGCTCCATTGT GCGACCTGTGAGAAGCCTCTAGAAGGGATTCTCTCCTCCTCTGCTTGGCACTTCACAC ACTCCCATTTACAATGCCACCATCTATGAAAATTCTTCTCCCAAGACCTATGTGGAGAG CTTCGAGAAAATGGGCATCTACCTCGCGGAGCCACAGTGGGCAGTGAGGTACCGGATC ATCTCTGGGGATGTGGCCAATGTATTTAAACTGAGGAGTATGTGGTGGGCAACTTCT GCTTCCTAAGAATAAGGACAAAGAGCAGCAACACAGCTCTTCTGAACAGAGAGGTGCG AGACAGCTACACCCCTCATCATCCAAGCCACAGAGAAGACCTTGGAGTTGGAAGCTTTG ACCCGTGTGGTGGTCCACATCCTGGACCAGAATGACCTGAAGCCTCTCTTCTCTCCAC CTTCGTACAGAGTCACCATCTCTGAGGACATGCCCCTGAAGAGCCCCATCTGCAAGGT GACTGCCACAGATGCTGATCTAGGCCAGAATGCTGAGTCTATTATGCCTTTAACACA AGGTCAGAGATGTTTGCCATCCATCCCACCAGCGGTGTGGTCACTGTGGCTGGGAAGC TTAACGTACACCTGGCGAGGAAAGCATGAGCTCCAGGTGCTAGCTGTGGACCGCATGCG GAAAATCTCTGAGGGCAATGGGTTTGGCAGCCTGGCTGCACTTGTGGTTTATGTGGAG CCTGCCCTCAGGAAGCCCCCAGCCATTGCTTCGGTGGTGGTGACTCCACCAGACAGCA ATGATGGTACCACCTATGCCACTGTACTGGTCGATGCAAAATAGCTCAGGAGCTGAAGT GGAGTCAGTGGAAAGTTGTTGGTGGTGACCCTGGAAAGCACTTCAAAGCCATCAAGTCT TATGCCCGGAGCAATGAGTTTCAAGTTTGGTGTCTGTCAAAGACATCAACTGGATGGAGT ACCTTCATGGGTTCAACCTCAGCCTCCAGGCCAGGAGTGGGAGCGGCCCTTATTTTTTA TTCCCAGATCAGGGGCTTTTACCTACCACCTTCCAAACTGTCTCCCTCAAATTCGAG AAGGCTGTTTACAGAGTGCAGCTTAGTGAGTTTTCCCTCCTGGCAGCCGCGTGGTGA TGGTGAGAGTCACCCAGCCTTCCCCAACCTGCAGTATGTTCTAAAGCCATCTTCAGA GAATGTAGGATTTAACTTAATGCTCGAACTGGGTTGATCACCACCACAAAGCTCATG GACTTCCACGACAGAGCCCACTATCAGCTACACATCAGAACCTCACCAGGCGAGGCCT CCACCGTGGTGGTCATTGACATTGTGGACTGCAACAACCATGCCCCCTCTTCAACAG GTCTTCCTATGATGGTACCTTGGATGAGAACATCCCTCCAGGCACCACTGTTTTGGCT GTGACTGCCACTGACCGGGATCATGGGGAAAATGGATATGTCACCTATTCCATTGCTG GACCAAAAGCTTTGCCATTTTCTATTGACCCCTACCTGGGGATCATCTCCACCTCCAA ACCCATGGACTATGAACTCATGAAAAGAATTTATACCTTCCGGGTAAAGAGCATCAGAC TGGGGATCCCCCTTTTCGCCGGGAGAAGGAAGTGCCATTTTCTTTCAGCTCAGGAAC TGAATGACAACCAGCCTATGTTTGAAGAAGTCAACTGTACAGGGTCTATCCGCCAAGA CTGGCCAGTAGGGAATCGATAATGACTATGTCAGCCATAGATGTGGATGAGCTTCAG AACCTAAAATACGAGATTGTATCAGGCAATGAACTAGAGTATTTTGATCTAAATCATT TCTCCGGAGTGATATCCCTCAAACGCCCTTTTATCAATCTTACTGCTGGTCAACCCAC CAGTTATTCCCTGAAGATTACAGCCTCAGATGGCAAAAACATGCTTACCCACCAACT TTGAATATTACTGTGGTGAAGGACCCTCATTTTGAAGTTCCTGTAACATGTGATAAAA CAGGGGTATTGACACAATTCACAAAGACTATCCTCCACTTTATTGGGCTTCAGAACCA GGAGTCCAGTGATGAGGAATTCATTCTTTAAGCACATATCAGATTAATCATTACACC CCACAGTTTGAAGACCACTTCCCCCAATCCATTGATGTCTTGAGAGTGTCCCTATCA ACACCCCTTGGCCCGCCTAGCAGCCACTGACCCTGATGCTGGTTTTAATGGCAAAC GGTCTATGTGATTGCAGATGGCAATGAGGAGGGCTGCTTTGACATAGAGCTGGAGACA GGGCTGCTCACTGTAGCTGCTCCCTTGGACTATGAAGCCACCAATTTCTACATCCTCA ATGTAACAGTATATGACCTGGGCACACCCAGAGTCTCTGGAAGCTGCTGACAGT GAATGTGAAAGACTGGAATGACAACGCACCCAGATTTCTCCCGGTGGGTACCAAGTTA ACCATCTCGGAGGACACAGAAGTTGGAACCAACAATTGCAGAGCTGACAACCAAGATG CTGACTCGGAAGACAATGGCAGGGTTGCTACACCCTGCTAAGTCCACAGAGAAGTT CTCCCTCCACCTCTCACTGGGGAAGTGGTTGTTACAGGACACCTGGACCGCGAATCA GAGCCTCGGTACATACTCAAGGTGGAGGCCAGGGATCAGCCAGCAAAGGCCACCAGC	

TCTTCTCTGTCACTGACCTGATAATCACATTGGAGGATGTCAACGACAACCTCTCCCCA GTGCATCACAGAACACAACAGGCTGAAGGTTCCAGAGGACCTGCCCCCGGGACTGTC TTGACATTTCTGGATGCCTCTGATCCTGACCTGGGCCCCGAGGTGAAGTGCGATATG TTCTGATGGATGGCGCCCATGGGACCTTCCGGGTGGACCTGATGACAGGGGCGCTCAT TCTGGAGAGAGAGCTGGACTTTGAGAGGCGAGCTGGGTACAATCTGAGCCTGTGGGCC AGTGATGGTGGGAGGCCCCCTAGCCCGCAGGACTCTCTGCCATGTGGAGGTGATCGTCC TGGATGTGAATGAGAATCTCCACCCTCCCCACTTTGCCTCCTTCGTGCACCAAGGCCA GGTGCAGGAGAACAGCCCCCTCGGGAACCTCAGGTGATTGTAGTGGCTGCCAGGACGAT GACAGTGGCTTGGATGGGGAGCTCCAGTACTTCTGCGTGCTGGCACTGGACTCGCAG CCTTCAGCATCAACCAAGATACAGGAATGATTACAGACTCTGGCACCCCTGGACCGAGA ATTTGCATCTTACTACTGGTTGACGGTATTAGCAGTGGACAGGGGTTCTGTGCCCCCTC TCTTCTGTAACCTGAAGTCTACATCGAGGTTACGGATGCCAATGACAACCCACCCCGA TGTCCCAAGCTGTGTTCTACCCCTCCATCCAGGAGGATGCTCCCGTGGGCACCTCTGT GCTTCAACTGGATGCCTGGGACCCAGACTCCAGCTCCAAAGGGAAGCTGACCTTCAAC ATCACCAGTGGGAACTACATGGGATTCTTTATGATTACCCCTGTTACAGGTCTCCTAT CTACAGCCCAGCAGCTGGACAGAGAGAACAAGGATGAACACATCCTGGAGGTGACTGT GCTGGACAATGGGGAACCCCTCACTGAAGTCCACCTCCAGGGTGGTGGTAGGCATCTTG GACGTCAATGACAATCCACCTATATTCTCCACAAGCTCTTCAATGTCCGCCCTTCCAG AGAGGCTGAGCCCTGTGTCCCCTGGGCCTGTGTACAGGCTGGTGGCTTCAGACCTGGA TGAGGGTCTTAATGGCAGAGTCACCTACAGTATCGAGGACAGCTATGAGGAGGCCTTC AGTATCGACCTGGTCACAGGTGTGGTTTCATCCAACAGCACTTTTACAGCTGGAGAGT ACAACATCCTAACGATCAAGGCAACAGACAGTGGGCAGCCACCACTCTCAGCCAGTGT CCGGCTACACATTGAGTGGATCCCTTGGCCCCCGGCCGTCTCCATCCCTCTGGCCTTT GATGAGACCTACTACAGCTTTACGGTCATGGAGACGGACCCCTGTGAACCACATGGTGG GGGTCATCAGCGTAGAGGGCAGACCCCGGACTCTTCTGGTTCAACATCTCAGGTGGGGA TAAGGACATGGACTTTGACATTGAGAAGACCACAGGCAGCATCGTCATTGCCAGGCCT CTTGATACCAGGAGAAGGTGCAACTATAACTTGACTGTTGAGGTGACAGATGGGTCCC GCACCATTGCCACACAGGTCCACATCTTCATGATTGCCAACATTAACCACCATCGGCC CCAGTTTCTGGAACTCGTTATGAAGTCAGAGTTCCCCAGGACACCGTGCCAGGGGTA GAGCTCCTGCGAGTCCAGGCCATAGATCAAGACAAGGGCAAAAGCCTCATCTATACCA TACATGGCAGCCAAGACCCAGGAAGTGCCAGCCTCTTCCAGCTGGACCCAAGCAGTGG TGTCCTGGTAACGGTGGGAAAATTGGACCTCGGCTCGGGGCCCTCCAGCACACATG ACAGTCATGGTCCGAGACCAGGAATACCTATCAAGAGGAACCTCGTGTGGGTGACCA TTCATGTGGAGGATGGAACCTCCACCCACCCCGCTTCACTCAGCTCCATTATGAGGC AAGTGTTCTTGACACCATAGCCCCCGGCACAGAGCTGCTGCAGGTCCGAGCCATGGAT GCTGACCGGGGAGTCAATGCTGAGGTCCACTACTCCCTCCTGAAAGGGAACAGCGAAG GTTTCTTCAACATCAATGCCCTGCTAGGCATCATTACTCTAGCTCAAAAGCTTGATCA GGCAAATCATGCCCCACATACTCTGACAGTGAAGGCAGAAGATCAAGGCTCCCCACAA TGGCATGACCTGGCTACAGTGATCATTATGATGCTATCCCTCAGATAGGAGTGCCCCCA TCTTTTCAAAATCTGAGTACTTTGTAGAGATCCCTGAATCAATCCCTGTTGGTTCCCC AATCCTCCTTGTCTCTGCTATGAGCCCCCTCTGAAGTTACCTATGAGTTAAGAGAGGGA AATAAGGATGGAGTCTTCTCTATGAACCTCATATTCTGGCCTTATTTCCACCGATGAGA AATTGGACCATGAGAAAATCTCGTCTTACCAGCTGAAAATCCGAGGCACCAATATGGC AGGTGCATTTACTGATGTATGGTGGTGGTTGACATAATTGATGAAAATGACAATGCT CCTATGTTCTTAAAGTCAACTTTTGTGGGCCAAATTAGTGAAAGCAGCTCCACTGTATA GCATGATCATGGATAAAAACAACACCCCTTTGTGATTATGCCTCTGACAGTGACAA AGAAGCTAATTCTTGTGGTCTATAAAATTTTGGAGCCGGAGGCCTTGAAGTTTTTC AAAATTGATCCCAGCATGGGAACCCTAACCATTGTATCAGAGATGGATTATGAGAGCA TGCCCTCTTTCCAATTCTGTGTCTATGTCCATGACCAAGGAAGCCCTGTATTATTTGC ACCCAGACCTGCCCAAGTCATCATTATGTGATGAGAGATGTGAATGATTCCCTCCCAGA TTCTCAGAACAGATATATGAGGTAGCAATAGTCGGGCCTATCCATCCAGGCATGGAGC TTCTCATGGTGGCGGCCAGCGATGAAGACTCAGAAGTCAATTATAGCAATCAAACTGG CAATGCTGATGAAGCTGTTACCATCCATCCTGTCACTGGTAGCATATCTGTGCTGAAT CCTGCTTTCTGGGACTCTCTCGGAAGCTCACCATCAGGGCTTCTGATGGCTTGTATC AAGACACTGCGCTGGTAAAAATTTCTTTGACCCAAGTGCTTGACAAAAGCTTGCACTT TGATCAGGATGTCTACTGGGCAGCTGTGAAGGAGAACTTGCAGGACAGAAAGGCACTG GTGATTCTTGGTGCCAGGGCAATCATTTGAATGACACCCCTTTCTACTTTCTCTTGA ATGGCACAGATATGTTTCATATGGTCCAGTCAGCAGGTGTGTTGCAGACAAGAGGTGT GGCGTTTGACCGGGAGCAGCAGGACACTCATGAGTTGGCAGTGGAAGTGAGGGACAAT CGGACACCTCAGCGGGTGGCTCAGGGTTTGGTCAGAGTCTTATGAGGATGTCAATG ACAATCCCCCAAATTTAAGCATCTGCCCTATTACACAATCATCCAAGATGGCACAGA
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

GCCAGGGGATGTCCTCTTTTCAGGTATCTGCCACTGATGAGGACTTGGGGACAAATGGG  
 GCTGTTACATATGAATTTGCAGAAGATTACACATATTTCCGAATTGACCCCTATCTTG  
 GGGACATATCACTCAAGAAACCCCTTTGATTATCAAGCTTTAAATAAATATCACCTCAA  
 AGTCATTGCTCGGGATGGAGGAACGCCATCCCTCCAGAGTGAGGAAGAGGTACTTGTC  
 ACTGTGAGAAATAAATCCAACCCACTGTTTCAGAGTCCTTATTACAAAGTCAGAGTAC  
 CTGAAAATATCACCCCTCTATACCCCAATTCTCCACACCCAGGCCCGAGTCCAGAGGG  
 ACTCCGGCTCATCTACAACATTGTGGAGGAAGAACCCTTGATGCTGTTCCACCACTGAC  
 TTCAAGACTGGTGTCTTAACAGTAACAGGGCCTTTGGACTATGAGTCCAAGACCAAAC  
 ATGTGTTACAGTCAGAGCCACGGATACAGCTCTGGGGTCATTTTCTGAAGCCACAGT  
 GGAAGTCCTAGTGGAGGATGTCAATGATAACCCTCCCACTTTTTCCCAATTGGTCTAT  
 ACCACTTCCATCTCAGAAGGCTTGCTGCTCAGACCCCTGTGATCCAAC'TGTTGGCTT  
 CTGACCAGGACTCAGGGCGGAACCGTGACGTCTCTTATCAGATTGTGGAGGATGGCTC  
 AGATGTTTCCAAGTTC'TTCCAGATCAATGGGAGCACAGGGGAGATGTCCACAGTTCAA  
 GAACTGGATTATGAAGCCCAACAACACTTTCATGTGAAAGTCAGGGCCATGGATAAAG  
 GAGATCCCCCACTCACTGGTGAAACCCTTGTGGTGTCAATGTGTCTGATATCAATGA  
 CAACCCCCCAGAGTTCAGACAACCTCAATATGAAGCCAATGTCAAGTGAAGTGGCAACC  
 TGTGGACACCTGGTTC'TTAAAGTCCAGGCTATTGACCCTGACAGCAGAGACACCTCCC  
 GCCTGGAGTACCTGATTCTTTCTGGCAATCAGGACAGGCAC'TTCTTCATTAACAGCTC  
 ATCGGGAATAATTTCTATGTTCAACCTTTGCAAAAAGCACCTGGACTCTTCTTACAAT  
 TTGAGGGTAGGTGCTTCTGATGGAGTCTTCCGAGCAACTGTGCCTGTGTACATCAACA  
 CTACAAATGCCAACAAGTACAGCCCAGAGTTCAGCAGCACCTTTATGAGGCAGAATT  
 AGCAGAGAATGCAATGGTGTGAACCAAGGTGATTGATTGTCTAGCCATAGACAAAGAT  
 AGTGGTCCCTATGGCACTATAGATTATACTATCATCAATAAACTAGCAAGTGAGAAGT  
 TCTCCATAAACCCCAATGGCCAGATTGCCACTCTGCAGAAACTGGATCGGGAAAATTC  
 AACAGAGAGAGTCATTGCTATTAAAGTCAATGGCTCGGGATGGAGGAGGAAGAGTAGCC  
 TTCTGCACGGTGAAGATCATCCTCACAGATGAAAATGACAACCCCCCACAGTTCAAAG  
 CATCTGAGTACACAGTATCCATTCAATCCAATGTCAAGTAAAGACTCTCCGGTTATCCA  
 GGTGTTGGCCTATGATGCAGATGAAGGTGAGAACGCAGATGTCACCTACTCAGTGAAC  
 CCAGAGGACCTAGTTAAAGATGTCATTGAAATTAACCCAGTCACTGGTGTGGTCAAGG  
 TGAAAGACAGCCTGGTGGGATTGGAAAATCAGACCCCTTGACTTCTTCATCAAAGCCCA  
 AGATGGAGGCCCTCCTCACTGGAACCTCTCTGGTGCCAGTACGACTTCAGGTGGTTCTCT  
 AAAAAAGTATCCTTACCGAAATTTTCTGAACCTTTGTATACTTTCTCTGCACCTGAAG  
 ACCTTCCAGAGGGGTCTGAAATTGGGATTGTTAAAGCAGTGGCAGCTCAAGATCCAGT  
 CATCTACAGTCTAGTGCAGGGCACTACACCTGAGAGCAACAAGGATGGTGTCTTCTCC  
 CTAGACCCAGACACAGGGGTCTATAAAGGTGAGGAAGCCCATGGACCACGAATCCACCA  
 AATTGTACCAGATTGATGTGATGGCACATTGCCTTCAGAACACTGATGTGGTGTCTCTT  
 GGTCTCTGTCAACATCCAAGTGGGAGACGTCAATGACAATAGGCCTGTATTTGAGGCT  
 GATCCATATAAGGCTGTCTCACTGAGAATATGCCAGTGGGGACCTCAGTCATTCAAG  
 TGACTGCCATTGACAAGGACACTGGGAGAGATGGCCAGGTGAGCTACAGGCTGTCTGC  
 AGACCCCTGGTAGCAATGTCCATGAGCTCTTTGCCATTGACAGTGAGAGTGGTGGATC  
 ACCACACTCCAGGAACCTTGACTGTGAGACCTGCCAGACTTATCATTTTTCATGTGGTGG  
 CCTATGACCACGGACAGACCATCCAGCTATCCTCTCAGGCCCTGGTTCCAGGTCTCCAT  
 TACAGATGAGAATGACAATGCTCCCCGATTGTCTTCTGAAGAGTACAGAGGATCTGTG  
 GTTGAGAACAGTGAGCCTGGCGAACTGGTGGCGACTCTAAAGACCCTGGATGCTGACA  
 TTTCTGAGCAGAACAGGCAGGTACCTGCTACATCACAGAGGGAGACCCCTGGGCCA  
 GTTTGGCATCAGCCAAGTTGGAGATGAGTGGAGGATTTCTCAAGGAAGACCCTGGAC  
 CGCGAGCATAACAGCCAAGTACTTGCTCAGAGTCACAGCATCTGATGGCAAGTTCCAGG  
 CTTCGGTCACTGTGGAGATCTTTGTCTTGACGTCAATGATAACAGCCACAGTGTTT  
 ACAGCTTCTCTATACTGGCAAGGTTTCATGAAGATGTATTTCCAGGACACTTCATTTTG  
 AAGGTTTCTGCCACAGACTTGACACTGATACCAATGCTCAGATCACATATTCTCTGC  
 ATGGCCCTGGGGCGCATGAATTCAGCTGGATCCTCATAACAGGGGAGCTGACCACACT  
 CACTGCCCTAGACCGAGAAAGGAAGGATGTGTTCAACCTTGTGGCAAGGCGACGGAT  
 GGAGGTGGCCGATCGTGCCAGGCAGACATCACCTCCATGTGGAGGATGTGAATGACA  
 ATGCCCCGCGGTTCTTCCCCAGCCACTGTGCTGTGGCTGTCTTCGACAACACCACAGT  
 GAAGACCCCTGTGGCTGTAGTATTTGCCCCGGGATCCCGACCAAGGCGCCAATGCCAG  
 GTGGTTTACTCTCTGCCGATTGAGCCGAAGGCCACTTTTCCATCGACGCCACCACGG  
 GGGTGATCCGCCTGGAAAAGCCGCTGCAGGTGAGGCCCCAGGCACCACTGGAGCTCAC  
 GGTCCGTGCCTCTGACCTGGGCACCCCAATACCGCTGTCCACGCTGGGCACCGTCACA  
 GTCTCGGTGGTGGGCCTAGAAGACTACCTGCCCCGTGTTCTGAACACCGGCACAGCG  
 TGCAGGTGCCCCGAGGACGCCCCACCTGGCACGGAGGTGCTGCAGCTGGCCACCTCAC  
 TCGCCGGGCGCAGAGAAGACCGCTACCGCGTGGTTCAGCGGGAACGAGCAAGGCAGG

TTCCGCCTGGATGCTCGCACAGGGATCCTGTATGTCAACGCAAGCCTGGACTTTGAGA  
 CAAGCCCCAAGTACTTCCTGTCCATTGAGTGCAGCCGGAAGAGCTCCTCTCCCTCAG  
 TGACGTGACCACAGTCATGGTCAACATCACTGATGTCAATGAACACCGGCCCAATTC  
 CCCCAGATCCATATAGCACAAAGGTCTTAGAGAATGCCCTTGTGGGTGACGTCATCC  
 TCACGGTATCAGCGACTGATGAAGATGGACCCCTAAATAGTGACATTACCTATAGCCT  
 CATAGGAGGGAACCAGCTTGGGCACTTCACCATTACCCCCAAAAGGGGGAGCTACAG  
 GTGGCCAAGGCCCTGGACCGGGAACAGGCCTCTAGTTATTCCCTGAAGCTCCGAGCCA  
 CAGACAGTGGGCAGCCTCCACTGCATGAGGACACAGACATCGCTATCCAAGTGGCTGA  
 TGTCAATGATAACCCACCGAGATTCTTCCAGCTCAACTACAGCACCACCTGTCCAGGAG  
 AACTCCCCCATTTGGCAGCAAAGTCCTGCAGCTGATCCTGAGTGACCCAGATTCTCCAG  
 AGAATGGCCCCCCTACTCGTTTTCGAATCACCAAGGGGAACAACGGCTCTGCCTTCCG  
 AGTGACCCCGGATGGATGGCTGGTGACTGCTGAGGGCCTAAGCAGGAGGGCTCAGGAA  
 TGGTATCAGCTTCAGATCCAGGCGTCAGACAGTGGCATCCCTCCCCTCTCGTCTTTGA  
 CGTCTGTCCGTGTCCATGTACAGAGCAGAGCCACTATGCACCTTCTGCTCTCCCACT  
 GGAGATCTTCATCACTGTTGGAGAGGATGAGTTCCAGGGTGGCATGGTGGGTAAGATC  
 CATGCCACAGACCGAGACCCCCAGGACACGCTGACCTATAGCCTGGCAGAAGAGGAGA  
 CCCTGGGCAGGCACCTTCTCAGTGGGTGCGCCTGATGGCAAGATTATCGCCGCCCAGGG  
 CCTGCCTCGTGGCCACTACTCGTTCAACGTCACGGTCAGCGATGGGACCTTCACCACG  
 ACTGCTGGGGTCCATGTGTACGTGTGGCATGTGGGGCAGGAGGCTCTGCAGCAGGCCA  
 TGTGGATGGGCTTCTACCAGCTCACCCCCAGGAGCTGGTGAGTGACCACTGGCGGAA  
 CCTGCAGAGGTTCCCTCAGCCATAAGCTGGACATCAAACGGGCTAACATTCACTTGGCC  
 AGCCTCCAGCCTGCAGAGGCCGTGGCTGGTGTGGATGTGCTCCTGGTCTTTGAGGGGC  
 ATTCTGGAACCTTCTACGAGTTTCAGGAGCTAGCATCCATCATCACTCACTCAGCCAA  
 GGAGATGGAGCATTCACTGGGGGTTGAGATGCGGTGAGCTATGCCCATGGTGGCCTGC  
 CAGGGGCCAACCTGCCAGGGTCAAATCTGCCATAACACAGTGCATCTGGACCCCAAGG  
 TTGGGGCCACGTACAGCACCGCCAGGCTCAGCATCCTAACCCCGCGGCACCACCTGCA  
 GAGGAGCTGCTCCTGCAATGGTACTGCTACAAGGTTCACTGGTCAGAGCTATGTGCGG  
 TACAGGGCCCCAGCGGCTCGGAACCTGGCACATCCATTCTATCTGAAAACACTCCAGC  
 CACAGGCCATTCTTCTATTACCAATGAAACAGCGTCCGTCTCCCTGAAGCTGGCCAG  
 TGGAGTGCCCCAGCTGGAATACCACTGTCTGGGTGGTCTTCTATGGAAACCTTTCTCTC  
 CAGCGCCATGTGAATGACCACGAGTGGCACTCCATCCTGGTGGAGGAGATGGACGCTT  
 CCATTGCGCTGATGGTTGACAGCATGGGCAACACCTCCCTTGTGGTCCAGAGAACTG  
 CCGTGGTCTGAGGCCCCGAAAGGCACCTCTTGCTGGGCGGCCTCATTCTGTTGCATTCT  
 TCCTCGAATGTCTCCAGGGCTTTGAAGGCTGCCTGGATGCTGTCTGTGGTCAACGAAG  
 AGGCTCTAGATCTGCTGGCCCTGGCAAGACGGTGGCAGGCTTGCTGAGAGACACAAGC  
 CCTACCCAGTGTGCTGCCTCCACAGTGACTACTGCAGCCAGAACACATGCCTCAATGGT  
 GGGAAGTGCTCATGGACCCATGGGGCAGGCTATGTCTGCAAATGTCCCCACAGTTCT  
 CTGGGAAGCACTGTGAACAAGGAAGGGAGAAGTGTACTTTTGCACCCTGCCTGGAAGG  
 TGGAACCTGCATCCTCTCCCCAAAGGAGCTTCTGTAACTGCCCTCATCCTTACACA  
 GGAGACAGGTGTGAATGGAGGCGAGGGGTTGTTTCAGAAGGACACTGCCTAGTCACTC  
 CCGAGATCCAAAGGGGGGACTGGGGGCAGCAGGAGTTACTGATCATCACAGTGGCCGT  
 GCGTTTCAATTATCATAAGCACTGTGCGGCTTCTCTTCTACTGCCGCCGTTGCAAGTCT  
 CACAAGCCTGTGGCCATGGAGGACCCAGACCTCCTGGCCAGGAGTGTGGTGTGTGACA  
 CCCAAGCCATGCCTGCCATCGAGCTCAACCCATTGAGTGCCAGCTCTTGAACAACCT  
 CAACCAACCGGAACCCAGCAAGGCCCTCTGTTCCAAATGAACTCGTCACATTGTGACCC  
 AATTCTAAGCAACGGCCAGTGGTCTGCAGTGTGCCCCCAGACTCCCGCCAGCTGCGG  
 TCCCTTCCCCTCTGACAATGAGCCTGTCAATTAAGAGAACCTGGTCCAGCGAGGAGAT  
 GGTGTACCCTGGCGGAGCCATGGTCTGGCCCCCTACTTACTCCAGGAACGAACGCTGG  
 GAATACCCCCACTCCGAAGTGAAGTCAAGGGCCCTCTGCCGCCCTCGGCTCACCGCCACT  
 CAACCCAGTCTGATGCCAGAGCCTAATGGCCTCTATGGGGGCTTCCCTTCCCTCCT  
 GGAGATGGAAAACAAGCGGGCACCTCTCCACCCCCGTTACAGCAACCAGAACCTGGAA  
 GATCTGATGCCCTCTCGGCCCTTAGTCCCCGGGAGCGCCTGGTTGCCCCCTGTCTCA  
 ATGAGTACACGGCCATCAGCTACTACCACTCGCAGTTCGGCAGGGAGGGGAGGGCC  
 CTGCTTGGCAGACGGGGCTACAAGGGGGTGGGTATGCGCCTCAGCCGAGCTGGGCCC  
 TCTTATGCTGTCTGTGAGGTGGAGGGGGCACCTCTTGCAGGCCAGGGCCAGCCCCGGG  
 TGCCCCCAACTATGAGGGCTCTGACATGGTGGAGAGTGATTATGGCAGCTGTGAGGA  
 GGTCAATGTTCTAGCTTCCATTCCCAGAGCAAGGCAGGCGGGAGGCCAAGGACTGGAC  
 TTGGCTTATTTCTTCTGTCTCGTAGGGGGTGAGTTGAGTGTGGCTGGGAGAGTGGGA  
 GGGAGGCCCTCAGCCAGGCTGTTGTCCCTTGAAATGTGCTCTTCCAATCCCCACCT  
 AGTCCCTGAGGGTGGAGGGAAGCTGAGGATAGAGCTCCAGAAACAGCACTAGGGTCCC  
 AGGAGAGGGGCATTTCTAGAGCAGTGACCTTGAAAACAGGAACAATTGACTCCTGG

	<p>GGTGGGCGACAGACAGGAGGGCTCCCTGATCTGCCGGCTCTCAGTCCCCGGGGCAAAG  CCTGATTGACTGTGCTGGCTCAACTTCACCAAGATGCATTCTCATACCTGCCACAGC  TCCATTTTGGAGGCAGGCAGGTTGGTGCCTGACAGACAACCCTACGCGGGCCGTACA  GAGGAGCTCTAGAGGGCTGCGTGGCATCCTCCTAGGGGCTGAGAGGTGAGCAGCAGGG  GAGCGGGCACAGTCCCCTCTGCCCCTGCCTCAGTCGAGCACTCACTGTGTCTTTGTCA  AGTGTCTGCTCCACGTCAAGGCACTGTGCTTTGCACCGGGGAGAAAATGGTGATGGAGG  GCAACAAGGACTCCGAGGAGCACCACCAGGCCTCGGGCCCCAGAGGTCCCCTCCTCA  GCCTACACGCAGAGGAACGGGGCCACCTCAGAGTCACACCCTGGCTGCCAGTCAGGG  CCTGCCAGGAGTCTACACAGCTCTGAACCTTCTTTGTTAAAGAATTAGACCTCATGG  AACTCTGGGTTCTTTCATCCCAAGTTTCCCAGGCACTTTTGGCCAAAGGAAGGAAGGAA  CTAATTCTTTCATTTTAAAAATTCTTAGGCACCTTTTGCACCTGTCTGTCTGGATGAGTT  TCCTCAATGGGATTTTTCTTCCCTAGACACAAGGAAGTCTGAACCTCTATTTAGGGCC  GGTTGGAAGCAGGGAGCTGGACCGCAGTGTCCAGGCTGGACACCTGCCATTGCCTCCT  CTCCACTGCAGACGCCTGCCATCAAGTATTACCTGCAGCGACTCAACCCTATGCATG  GAGGGTCAATGTGGGCACATGTCTACACATGTGGGTGCCCATGGATAGTACGTGTGTA  CACATGTGTAGAGTGTATGTAGCCAGGAGTGGTGGGGACCAGAAGCCTCTGTGGCCTT  TGGTGACCTCACCCTCCCTCCCACCCAGTCCCTCCCTCTGGTCCACTGCCTTTTCAT  ATGTGTTGTTTCTGGAGACAGAAGTCAAAAGGAAGAGCAGTGGAGCCTTGCCACAGG  GCTGCTGCTTCATGCGAGAGGGAGATGTGTGGGCGAGAGCCAATTTGTGTGAGTGGTT  TGTGGCTGTGTGTGACTGTGAGTGTGAGTGACAGATACATAGTTTCATTGGTTCATT  TTTTTTTTTAACAATAAAGTATCTTTTTTTTACTGTT</p>
	<p>ORF Start: ATG at 14      ORF Stop: TAG at 13061</p>
	<p>SEQ ID NO: 112      4349 aa      MW at 479387.3 Da</p>
NOV28a, CG51923-01 Protein Sequence	<p>MTIALLGFAIFLLHCATCEKPLEGILSSSAWHFTSHYNNATIYENSSPKTYVESFEKM  GIYLAEPQWAVRYRIISGDVANVFKTEYVVGNFCLIRIRTKSSNTALLNREVRDSYT  LI IQATEKTLEALTRVVHILDQNDLKPLFSPPSYRVTI SEDMPLKSPICKVTATD  ADLGQNAEFYYAFNTRSEMFAIHPTSGVTVAGKLNVTWRGKHELQVLAVDRMRKISE  GNGFGSLAALVVHVEPALRKPPAIASVVVTPPDSNDGTTYATVLVDANSSGAEVESVE  VVGDPGKHFKAIKSYARSNEFSLVSVKIDINWMEYLHGFNLSQLARSGSGPYFYSQIR  GFHLPPSKLSSLFKFAVYRVQLSEFSPPGSRVVMVRVTPAFPNLQYVLKPSSENVGF  KLNARTGLITTTKLMDFHDRHYQLHIRTSPGQASTVVVIDIVDCNNHAPLFNRSSYD  GTLDENIPPGTSVLAVTATDRDHGENGYVTYSIAGPKALPFSIDPYLGIISTSKPMDY  ELMKRIYTRVRASDWGSPFRREKEVSI FLQLRNLNDNQPMFEVNCITGSIRQDWPVG  KSIMTMSAIDVDELQNLKYEIVSGNELEYFDLNFHSGVISLKRPNINLTAGQPTSYSL  KITASDGKNYASPTTLNITVVKDPHFVEVPVTCDKTGVLTQFTKTI LHFIFGLQNESSD  EEFTSLSTYQINHYTPQFEDHFPQSIDVLESVPINTPLARLAATDPDAGFNGKLVYVI  ADGNEEGCFDIELETGLLTVAAPLDYEATNFYILNVTVYDLGTPQKSSWKLLTVNVKD  WNDNAPRFPPGGYQLTISEDTEVGTTIAELTTKBDASEDNGRVRYTLLSPTEKFSLHP  LTGELVVVTGHLDTRESEPRYILKVEARDQPSKGHQLFSVTDLIITLEDVNDNSPQCITE  HNRLKVPEDLP PGTVLTFDASDPDLGPAGEVRYVLM DGAHGTFRV DLMTGALILERE  LDFERRAGYNLSLWASDGGRLARRTLCHVEIVLDVNENLHPPHFASFVHQGVQEN  SPSGTQVIVVAAQDDDSGLDGLQYFLRAGTGLAAFSINQDTGMIQTLAPLDREFASY  YWLTVLAVDRGSVPLSSVTEVYIEVTDANDNPPQMSQAVFYPSIQEDAPVGTSVLQLD  AWDPDSSSKGKLT FNITSGNYMGFFMIHPVTGLLSTAQQLDRENKDEHILEVTVDNG  EPSLKSTSRVVGILDVNDNPPIFSHKLFNVRLPERLSPVSPGPVYRLVASDLDEGLN  GRVTYSIEDSYEEAFSIDLVTGVVSSNSTFTAGEYNILTIKATDSGQPPLSASVRLHI  EWIPWPRSSIPLAFDETYYSFTVMETDPVNHMVGVISVEGRPGLFWFNISGGDKDMD  FDIEKTTGSIV IARPLDTRRRSNYNLTVEVTDGSR TIATQVHIFMIANINHHRPQFLE  TRYEV RVPQDTPVGV ELLRVQAIDQDKGKSLIYTIHGSQDPGSASLFLQDPSSGVLVT  VGKLDLGSGPSQHTLTVMVRDQEIP IKRNFVWVTIHVEDGNLHPPRFTQLHYEASVPD  TIAPGTELLQVRAMDARGVNAEVHYSLLKGNSEGFNINALLGIITLAQLKDQANHA  PHTLTVKAEDQGSQWHDLATV IIVYPSDRSAPIFSKSEYFVEIPESIPVGSPI LLV  SAMPSPSEVTYELREGNKDGVFSMNSYSGLISTQKKLDHEKISSYQLKIRGSNMAGAFT  DVMVVVDI IDENDNAPMFLKSTFVGQISEAAPLYSMIMDKNNNP FVIHASDSDKEANS  LLVYKILEPEALKFFKIDPSMGTLTIVSEMDYESMPSFQFCVYVHDQGSPLVFAPRPA  QVI I HVRDVNDSPPRFSEQIYEVAIVGPIHPGMELLMVRASDEDSEVNYSIKTGNADE  AVTIHPVTGSI SVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQVLDKSLQFDQDV  YWAAVKENLQDRKALVILGAQGNHNDTLSYFLLNGTDMFHMVQSAGVLQTRGVAFDR  EQQDTHELAVEVRDNRTPQRVAQGLVRVSI EDVNDNPPKFKHLPYYTIIQDGTPEGDV</p>

	<p>LFQVSATDEDLGTNGAVTYEFAEDYTYFRIDPYLGDISLKKPFDYQALNKYHLKVIAR  DGGTPSLQSEEEVLVTVRNKSNPLFQSPYYKVRVPENITLYTPILHTQARSPEGLRLI  YNIVEEEPLMLFTTDFKTGVLTVTGPLDYESKTKHVFTVRATDTALGSFSEATVEVLV  EDVNDNPPTFSQLVYTTSISEGLPAQTPVIQLLASDQDSGRNRDVSQIVEDGSDVSK  FFQINGSTGEMSTVQELDYEAQQHFHVKVRAMDKGDPPLTGETLVVVNVSDINDNPPE  FRQPQYEANVSELATCGHLVLKVQAIDPDSRDTSRLEYLILSGNQDRHFFINSSSGII  SMFNLCKKHLDDSSYNLRVGASDGVFRATVPVYINTTNANKYSPEFQQHLYEAELEAENA  MVGTKVIDLLAIDKDSGPYGTIDYTIINKLASEKFSINPNGQIATLQKLDRENSTERV  IAIKVMARDGGGRVAFCTVKIILTDENDNPPQFKASEYTVSIQSNVSKDSPVIQVLAY  DADEGQNADVITYSVNPEDLVKDVEINPVTGVVKVSDLVGLENQTLDDFFIKAQDGGP  PHWNSLVPVRLQVVPKKVSLPKFSEPLYTFSAPEDLPEGSEIGIVKAVAAQDPVIYSL  VRGTTPESENKDGVSFLDPDTGVIKVRKPMDEHSTKLYQIDVMAHCLQNTDVSLSVSN  IQVGDVNDNRPVFEADPYKAVLTENMPVGTSVIQVTAIDKDTGRDGQVSRYLSADPGS  NVHELFAIDSESGWITTLQELDCETCQTYHFHVAYDHGQTIQLSSQALVQVSITDEN  DNAPRFASEEYRGSVENSEPGEIVATLTKTLDAIDISEQNQRVTCYITEGDPGQFGIS  QVGDEWRISSRKTLTREHTAKYLLRVITASDGKFQASVTVEIFVLDVNDNSPQCSQLLY  TGKVHEDVFPGHFILKVSATDLDTDTNAQITYSLHGPAGHEFKLDPHTGELTTLTALD  RERKDVFNLVAKATDGGGRSCQADITLHVEDVNDNAPRFFPSHCNAVAVFNTTVPV  AVVFARDPDQGANAVVYSLPDSAEGHFSIDATTGVIRLEKPLQVRPQAPLELTVRAS  DLGTPILPLSTLGTVTVSUVGLEDYLPVFLNTEHSVQVPEDAPPGTEVLQLATLTRPGA  EKTGYRVVSGNEQGRFRLDARTGILYVNASLDFETSPKYFLSIECSRKSSSSLSDVTT  VMVNITDVNEHRPQFPQDPYSTRVLENALVGDVILTVSATDEDDGPLNSDITYSLIGGN  QLGHFTIHPKKGELQVAKALDREQASSYSLKLRATDSGQPPLHEDTDIAIQVADVNDN  PFRFFQLNYSTTVQENSPIGSKVLQILSDPDSPENGPPYSFRITKGNNGSAFRVTPD  GWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTSVRVHVTEQSHYAPSALPLEIFI  TVGEDEFQGGMVGKIHATDRDPQDTLTYSLAEETLGRHFSVGAPDGKIIAAQGLPRG  HYSFNVTVSDGFTTTAGVHVYVHVQGEALQQAMWMGFYQLTPEELVSDHWRNLQRF  LSHKLDIKRANIHLASLQPAEAVAGVDVLLVFEGHSGTFYEFQELASIIITHSAKEMEH  SVGVQMRAMPMPVPCQGPTCQGQICHNTVHLDPKVGPITYSTARLSILTPRHHLQRSCS  CNGTATRFSGQSYVRYRAPAARNWHIHFYKTLQPPQAILFTNETASVSLKLASGVPO  LEYHCLGGFYGNLSSQRHVNDHEWHSILVEEMDASIRLMVDSMGNTSLVVPENCRLR  PERHLLGGLILLHSSSNVSQGFEGCLDAVVVNEEALDLLAPGKTVAGLLETQALTQC  CLHSDYCSQNTCLNGGKCSWTHGAGYVCKCPPQFSGKHCEQGRENCTFAPCLEGGTCI  LSPKGASCNCPPHYTGDRCEMEARGCSEGHCLVTPETIQRGDWGQQLIITVAVAFII  ISTVGLLFYCRCKSHKPVAMEDPDLARSVGVDTQAMPAIELNPLSASSCNNLNQPE  PSKASVPNELVTFGPNSKQRPVVCVPPRLPPAAVPSHSDNEPVIKRTWSSEEMVPG  GAMVWPPTYSRNERWEYPHSEVTQGPLPPSAHRHSTPVVMPENGLYGGFPFPLEMEN  KRAPLPPTYSNQNLLEDLMPSPRPPSPRERLVAPCLNEYTAISYYHSQFRQGGGGPCLAD  GGYKGVGMRLSRAGPSYAVCEVEGAPLAGQGQPRVPPNYEGSDMVESDYGSCEEVMF</p>
	<p>SEQ ID NO: 113</p> <p>14279 bp</p>
NOV28b, CG51923-03 DNA Sequence	<p>GGAGTTTTCCACCATGACTATTGCCCTGCTGGGTTTTGCCATATTCTTGCTCCATTGT  GCGACCTGTGAGAAGCCTCTAGAAGGGATTCTCTCCTCCTCTGCTTGGCACTTCACAC  ACTCCCATTAACAATGCCACCATCTATGAAAATTCTTCTCCCAAGACCTATGTGGAGAG  CTTCGAGAAAATGGGCCTTACCTCGCGGAGCCACAGTGGGCAGTGAGGTACCGGATC  ATCTCTGGGGATGTGGCCAATGTATTTAAAACTGAGGAGTATPTGTTGGGCAACTTCT  GCTTCCTAAGAATAAGGACAAAGAGCAGCAACACAGCTCTTCTGAACAGAGAGGTGCG  AGACAGCTACACCCTCATCATCCAAGCCACAGAGAAGACCTTGGAGTTGGAAGCTTTG  ACCCGTGTGGTGGTCCACATCCTGGACCAGAATGACCTGAAGCCTCTCTTCTCTCCAC  CTTCGTACAGAGTCACCATCTCTGAGGACATGCCCCGAAGAGCCCCATCTGCAAGGT  GACTGCCACAGATGCTGATCTAGGCCAGAATGCTGAGTTCTATTATGCCTTTAACACA  AGGTGAGAGATGTTTGCCATCCATCCCACAGCGGTGTGGTCACTGTGGCTGGGAAGC  TTAACGTCACCTGGCGAGGAAAGCATGAGCTCCAGGTGCTAGCTGTGGACCGCATGCG  GAAAATCTCTGAGGGCAATGGGTTTGGCAGCCTGGCTGCACTTGTGGTTTCATGTGGAG  CCTGCCCTCAGGAAGCCCCAGCCATTGCTTCGGTGGTGGTGACTCCACCAGCAGCA  ATGATGGTACCACCTATGCCACTGTACTGGTTCGATGCAAATAGCTCAGGAGCTGAAGT  GGAGTCAGTGAAGTTGTTGGTGGTGACCCTGGAAAGCACTTCAAAGCCATCAAGTCT  TATGCCCGGAGCAATGAGTTTCAAGTTTGGTGTCTGTCAAAGACATCAACTGGATGGAGT  ACCTTCATGGGTTCAACCTCAGCCTCCAGGCCAGGAGTGGGAGCGGCCCTTATTTTTA  TTCCAGATCAGGGGCTTTACCTACCACCTTCCAAACTGTCTTCCCTCAAATTCGAG</p>



AAGGCTGTTTACAGAGTGCAGCTTAGTGAGTTTTCCCTCCTGGCAGCCGCGTGGTGA  
TGGTGAGAGTCACCCAGCCTTCCCCAACCTGCAGTATGTTCTAAAGCCATCTTCAGA  
GAATGTAGGATTTAACTTAATGCTCGAACTGGGTTGATCACCACCACAAAGCTCATG  
GACTTCCACGACAGAGCCCACTATCAGCTACACATCAGAACCTCACCAGGCGCAGGCCT  
CCACCGTGGTGGTCATTGACATTGTGGACTGCAACAACCATGCCCCCTCTTCAACAG  
GTCTTCCTATGATGGTACCTTGGATGAGAACATCCCTCCAGGCACCAAGTGTGTTGGCT  
GTGACTGCCACTGACCGGGATCATGGGGAAAATGGATATGTCACCTATTCCATTGCTG  
GACCAAAAGCTTTGCCATTTTCTATTGACCCCTACCTGGGGATCATCTCCACCTCCAA  
ACCCATGGACTATGAACCTCATGAAAAGAATTTATACCTTCCGGGTAAAGAGCATCAGAC  
TGGGGATCCCCTTTTTCGCCGGGAGAAGTGTCCATTTTTCTCAGCTCAGGAACCT  
TGAATGACAACCAGCCTATGTTTGAAGAAGTCAACTGTACAGGGTCTATCCGCCAAGA  
CTGGCCAGTAGGGAAATCGATAATGACTATGTCAGCCATAGATGTGGATGAGCTTCAG  
AACCTAAAATACGAGATTGTATCAGGCAATGAACTAGAGTATTTTGATCTAAATCATT  
TCTCCGGAGTGATATCCCTCAAACGCCCTTTTATCAATCTTACTGCTGGTCAACCCAC  
CAGTTATTCCCTGAAGATTACAGCCTCAGATGGCAAAAACCTATGCCTCACCCACAACCT  
TTGAATATTACTGTGGTGAAGGACCCTCATTTTGAAGTTCCTGTAACATGTGATAAAA  
CAGGGGTATTGACACAATTCACAAAGACTATCCTCCACTTTATTGGGCTTCAGAACCA  
GGAGTCCAGTGATGAGGAATTCACCTTCTTTAAGCACATATCAGATTAATCATTACACC  
CCACAGTTTGAGGACCCTTCCCCCAATCCATTGATGTCTTGAGAGTGTCCCTATCA  
ACACCCCTTGGCCCGCCTAGCAGCCACTGACCCTGATGCTGGTTTTTAATGGCAAACCT  
GGTCTATGTGATTGCAGATGGCAATGAGGAGGGCTGCTTTGACATAGAGCTGGAGACA  
GGGCTGCTCACTGTAGCTGCTCCCTTGGACTATGAAGCCACCAATTTCTACATCCTCA  
ATGTAACAGTATATGACCTGGGCACACCCAGAAAGTCTCCTGGAAGCTGCTGACAGT  
GAATGTGAAAGACTGGAATGACAACGCACCCAGATTTCTCCCGGTGGGTACCAGTTA  
ACCATCTCGGAGGACACAGAAGTTGGAACCACAATTGCAGAGCTGACAACCAGATG  
CTGACTCGGAAGACAATGGCAGGGTTTCGTACACCCCTGCTAAGTCCACAGAGAAGTT  
CTCCCTCCACCCCTCACTGGGGAACCTGGTTGTTACAGGACACCTGGACCGCGAATCA  
GAGCCTCGGTACATACTCAAGGTGGAGGCCAGGGATCAGCCAGCAAAGGCCACAGC  
TCTTCTCTGTCACTGACCTGATAATCACATTGGAGGATGTCAACGACAACCTCTCCCA  
GTGCATCACAGAACACAACAGGCTGAAGGTTCCAGAGGACCTGCCCCCGGGACTGTC  
TTGACATTTCTGGATGCCTCTGATCCTGACCTGGGCCCCGAGGTGAAGTGCGATATG  
TTCTGATGGATGGCGCCCATGGGACCTTCCGGGTGGACCTGATGACAGGGGCGCTCAT  
TCTGGAGAGAGAGCTGGACTTTGAGAGGCGAGCTGGGTACAATCTGAGCCTGTGGGCC  
AGTGATGGTGGGAGGCCCTAGCCCGCAGGACTCTCTGCCATGTGGAGGTGATCGTCC  
TGGATGTGAATGAGAATCTCCACCCCTCCCCACTTTGCCTCCTTCGTGCACCAGGGCCA  
GGTGCAGGAGAACAGCCCCCTCGGGAACCTCAGGTGATTGTAGTGGCTGCCAGGACGAT  
GACAGTGGCTTGGATGGGGAGCTCCAGTACTTCTGCGTGTGGCACTGGACTCGCAG  
CCTTCAGCATCAACCAAGATACAGGAATGATTGAGACTCTGGCACCCCTGGACCGAGA  
ATTTGCATCTTACTACTGGTTGACGGTATTAGCAGTGGACAGGGGTTCTGTGCCCCCTC  
TCTTCTGTAACGAAGTCTACATCGAGGTTACGGATGCCAATGACAACCCACCCAGAA  
TGTCCCAAGCTGTGTTCTACCCCTCCATCCAGGAGGATGCTCCCGTGGGCACCTCTGT  
GCTTCAACTGGATGCCTGGGACCCAGACTCCAGCTCCAAAGGGAGCTGACCTTCAAC  
ATCACAGTGGGAACTACATGGGATTCTTTATGATTACCCCTGTTACAGGTCTCCTAT  
CTACAGCCCAGCAGCTGGACAGAGAGAAACAAGGATGAACACATCCTGGAGGTGACTGT  
GCTGGACAATGGGGAACCCTCACTGAAGTCCACCTCCAGGGTGGTGGTAGGCATCTTG  
GACGTCAATGACAATCCACCTATATTCTCCACAAGCTCTTCAATGTCCGCCCTTCCAG  
AGAGGCTGAGCCCTGTGTCCCCTGGGCCTGTGTACAGGCTGGTGGCTTCAGACCTGGA  
TGAGGGTCTTAATGGCAGAGTCACCTACAGTATCGAGGACAGCTATGAGGAGGCCTTC  
AGTATCGACCTGGTCACAGGTGTGGTTTCATCCAACAGCACTTTTACAGCTGGAGAGT  
ACAACATCCTAACGATCAAGGCAACAGACAGTGGGCAGCCACCACTCTCAGCCAGTGT  
CCGGCTACACATTGAGTGGATCCCTTGGCCCCGGCCGTCTCCATCCCTCTGGCCTTT  
GATGAGACCTACTACAGCTTTACGGTCATGGAGACGGACCTGTGAACCACATGGTGG  
GGGTCATCAGCGTAGAGGGCAGACCCGGACTCTTCTGGTTCAACATCTCAGGTGGGGA  
TAAGGACATGGACTTTTGACATTGAGAAGACCACAGGCAGCATCGTCATTGCCAGGCCT  
CTTGATAACCAGGAGAAGGTGCAACTATAAATTGACTGTTGAGGTGACAGATGGTGGTCCC  
GCACCATTGCCACACAGGTCCACATCTTCATGATTGCCAACATTAACCACCATCGGCC  
CCAGTTTCTGGAAACTCGTTATGAAGTCAGAGTTCCCCAGGACACCGTGCCAGGGGTA  
GAGCTCCTGCGAGTCCAGGCCATAGATCAAGACAAGGGCAAAAGCCTCATCTATACCA  
TACATGGCAGCCAAGACCCAGGAAGTGCCAGCCTCTTCCAGCTGGACCCAAGCAGTGG  
TGTCTTGGTAACGGTGGGAAAATTGGACCTCGGCTCGGGGCCCTCCAGCACACACTG  
ACAGTCATGGTCCGAGACCAGGAATACCTATCAAGAGGAACTTCGTGTGGGTGACCA

TTCATGTGGAGGATGGAAACCTCCACCCACCCCGCTTCACTCAGCTCCATTATGAGGC  
 AAGTGTTCCTGACACCATAGCCCCGGCACAGAGCTGCTGCAGGTCCGAGCCATGGAT  
 GCTGACCGGGGAGTCAATGCTGAGGTCCACTACTCCCTCCTGAAAGGGAACAGCGAAG  
 GTTTCTTCAACATCAATGCCCTGCTAGGCATCATTACTCTAGCTCAAAAGCTTGATCA  
 GGCAAAATCATGCCCCACATACTCTGACAGTGAAGGCAGAAGATCAAGGCTCCCCACAA  
 TGGCATGACCTGGCTACAGTGAATCATGTCTATCCCTCAGATAGGAGTGGCCCCCA  
 TCTTTTCAAAATCTGAGTACTTTGTAGAGATCCCTGAATCAATCCCTGTTGGTTCCCC  
 AATCCTCCTTGTCTCTGCTATGAGCCCCCTCTGAAGTTACCTATGAGTTAAGAGAGGGA  
 AATAAGGATGGAGTCTTCTCTATGAACATCATTTCTGGCCTTATTTCCACCCAGAAGA  
 AATTGGACCATGAGAAAATCTCGTCTTACCAGCTGAAAATCCGAGGCAGCAATATGGC  
 AGGTGCATTTACTGATGTCTGATGGTGGTGGTTGACATAATTGATGAAAATGACAATGCT  
 CCTATGTTCTTAAAGTCAACTTTTGTGGGCCAAATTAGTGAAAGCAGCTCCACTGTATA  
 GCATGATCATGGATAAAAAACAACCCCTTTGTGATTTCATGCCTCTGACAGTGACAA  
 AGAAGCTAATTCCTTGTGGTCTATAAAATTTTGGAGCCGGAGGCCTTGAAGTTTTC  
 AAAATTGATCCCAGCATGGGAACCCTAACCATTGTATCAGAGATGGATTATGAGAGCA  
 TGCCCTCTTTCCAATTCTGTGTCTATGTCCATGACCAAGGAAGCCCTGTATTATTTGC  
 ACCCAGACCTGCCCAAGTCATCATTCTATGTCTGAGAGATGTGAATGATTCCCCCTCCAGA  
 TTCTCAGAACAGATATATGAGGTAGCAATAGTCGGGCCTATCCATCCAGGCATGGAGC  
 TTCTCATGGTGCAGGGCCAGCGATGAAGACTCAGAAGTCAATTATAGCATCAAACTGG  
 CAATGCTGATGAAGCTGTTACCATCCATCCTGTCACTGGTAGCATATCTGTGCTGAAT  
 CCTGCTTTCCTGGGACTCTCTCGGAAGCTCACCATCAGGGCTTCTGATGGCTTGTATC  
 AAGACACTGCGCTGGTAAAAATTTCTTTGACCCAAGTGCTTGACAAAAGCTTGCAGTT  
 TGATCAGGATGTCTACTGGGCAGCTGTGAAGGAGAACTTGACAGGACAGAAAGGCACTG  
 GTGATTCTTGGTGGCCAGGGCAATCATTGTAATGACACCCCTTCTACTTTCTCTTGA  
 ATGGCACAGATATGTTTCATATGGTCCAGTCAGCAGGTGTGTTGCAGACAAGAGGTGT  
 GGCCTTTGACCGGGAGCAGCAGGACACTCATGAGTTGGCAGTGGAAGTGAGGGACAAT  
 CGGACACCTCAGCGGGTGGCTCAGGGTTTGGTCAGAGTCTCTATTGAGGATGTCAATG  
 ACAATCCCCCAAATTTAAGCATCTGCCCTATTACACAATCATCCAAGATGGCACAGA  
 GCCAGGGGATGTCTCTTTTCAGGTATCTGCCACTGATGAGGACTTGGGGACAAATGGG  
 GCTGTTACATATGAATTTGCAGAAGATTACACATATTTCCGAATTGACCCCTATCTTG  
 GGGACATATCACTCAAGAAACCCTTTGATTATCAAGCTTTAAATAAATATCACCTCAA  
 AGTCATTGCTCGGGATGGAGGAACGCCATCCCTCCAGAGTGAGGAAGAGGTACTTGTCT  
 ACTGTGAGAAATAAATCCAACCCACTGTTTCAGAGTCCCTATTACAAAGTCAGAGTAC  
 CTGAAAATATCACCCCTCTATACCCCAATTCTCCACACCCAGGCCGGAGTCCAGAGGG  
 ACTCCGGCTCATCTACAACATTGTGGAGGAAGAACCCTTGATGCTGTTCCACTGAC  
 TTCAAGACTGGTGTCTTAACAGTAACAGGGCCTTTGGACTATGAGTCCAAGACCAAAC  
 ATGTGTTACAGTCAGAGCCACGGATACAGCTCTGGGGTCATTTTCTGAAGCCACAGT  
 GGAAGTCTAGTGGAGGATGTCAATGATAACCCTCCCACTTTTCCCAATTGGTCTAT  
 ACCACTTCCATCTCAGAAGGCTTGCCTGCTCAGACCCCTGTGATCCAACTGTTGGCTT  
 CTGACCAGGACTCAGGGCGGAACCGTGACGTCTCTTATCAGATTGTGGAGGATGGCTC  
 AGATGTTTCCAAGTTCCTCCAGATCAATGGGAGCACAGGGGAGATGTCCACAGTTCAA  
 GAAGTGGATTATGAAGCCCAACAACACTTTTATGTGAAAGTCAGGGCCATGGATAAAG  
 GAGATCCCCCACTCACTGGTGAACCCCTTGTGGTTGTCAATGTGTCTGATATCAATGA  
 CAACCCCCCAGAGTTCAGACAACCTCAATATGAAGCCAATGTCAAGTGAAGTGGCAACC  
 TGTGGACACCTGGTTCCTTAAAGTCCAGGCTATTGACCCTGACAGCAGAGACACCTCCC  
 GCCTGGAGTACCTGATTCTTTCTGGCAATCAGGACAGGCACTTCTTCATTAACAGCTC  
 ATCGGGAATAATTTCTATGTTCAACCTTTGCAAAAAGCACCTGGACTCTTCTTACAAT  
 TTGAGGGTAGGTGCTTCTGATGGAGTCTTCCGAGCAACTGTGCCTGTGTACATCAACA  
 CTACAAATGCCAACAAAGTACAGCCCAGAGTTCAGCAGCACCTTTATGAGGCAGAATT  
 AGCAGAGAATGCAATGGTTGGAACCAAGGTGATTGATTGCTAGCCATAGACAAAGAT  
 AGTGGTCCCTATGGCACTATAGATTATACATCATCAATAAAGTACGAAGTGAGAAGT  
 TCTCCATAAAACCCCAATGGCCAGATTGCCACTCTGCAGAAACTGGATCGGGAAAATTC  
 AACAGAGAGAGTCATTGCTATTAAGGTCTATGGCTCGGGATGGAGGAGGAAGAGTAGCC  
 TTCTGCACGGTGAAGATCATCCTCAGAGATGAAAATGACAACCCCCCACAGTTCAAAG  
 CATCTGAGTACACAGTATCCATTCAATCCAATGTCAAGTAAAGACTCTCCGTTTATCCA  
 GGTGTTGGCCTATGATGCAGATGAAGGTGAGAACGCAGATGTCACTACTCAGTGAAC  
 CCAGAGGACCTAGTTAAAGATGTCAATTGAAATTAACCCAGTCACTGGTGTGGTCAAGG  
 TGAAAGACAGCCTGGTGGGATTGGAATTCAGACCCCTTGACTTCTTCATCAAAGCCCA  
 AGATGGAGGCCCTCCTCACTGGAACCTCTGTTGCCAGTACGACTTCAGTGGTTCCCT  
 AAAAAAGTATCCTTACCGAAATTTCTGAACCTTTGTATACTTCTCTGTCAGTGAAG  
 ACCTTCCAGAGGGGTCTGAAATTGGGATTGTTAAAGCAGTGGCAGCTCAAGATCCAGT

CATCTACAGTCTAGTGCGGGGCACTACACCTGAGAGCAACAAGGATGGTGTCTTCTCC  
CTAGACCCAGACACAGGGGTCATAAAGGTGAGGAAGCCCATGGACCAGGAATCCACCA  
AATTGTACCAGATTGATGTGATGGCACATTGCCTTCAGAACACTGATGTGGTGTCTT  
GGTCTCTGTCAACATCCAAGTGGGAGACGTCAATGACAATAGGCCTGTATTTGAGGCT  
GATCCATATAAGGCTGTCTCACTGAGAATATGCCAGTGGGGACCTCAGTCATTCAAG  
TGACTGCCATTGACAAGGACACTGGGAGAGATGGCCAGGTGAGCTACAGGCTGTCTGC  
AGACCCTGGTAGCAATGTCCATGAGCTCTTTGCCATTGACAGTGAGAGTGGTTGGATC  
ACCACACTCCAGGAACCTTGACTGTGAGACCTGCCAGACTTATCATTTTCATGTGGTGG  
CCTATGACCACGGACAGACCATCCAGCTATCCTCTCAGGCCCTGGTTCAGGTCTCCAT  
TACAGATGAGAATGACAATGCTCCCCGATTTGCTTCTGAAGAGTACAGAGGATCTGTG  
GTTGAGAACAGTGAGCCTGGCGAACTGGTGGCGACTCTAAAGACCCTGGATGCTGACA  
TTTCTGAGCAGAACAGGCAGGTACCTGCTACATCACAGAGGGAGACCCCTGGGCCA  
GTTTGGCATCAGCCAAGTTGGAGATGAGTGGAGGATTTCCCTCAAGGAAGACCCTGGAC  
CGCGAGCATACAGCCAAGTACTTGCTCAGAGTCACAGCATCTGATGGCAAGTTCAGG  
CTTCGGTCACTGTGGAGATCTTTGTCTGGACGTCAATGATAACAGCCCACAGTGTTC  
ACAGCTTCTCTATACTGGCAAGGTTTCATGAAGATGTATTTCCAGGACACTTCATTTTG  
AAGGTTTCTGCCACAGACTTGACACTGATACCAATGCTCAGATCACATATTTCTCTGC  
ATGGCCCTGGGGCGCATGAATTCAAGCTGGATCCTCATACAGGGGAGCTGACCACACT  
CACTGCCCTAGACCGAGAAAGGAAGGATGTGTTCAACCTTGTGTTGCCAAGGCGACGGAT  
GGAGGTGGCCGATCGTGCCAGGCAGACATCACCTCCATGTGGAGGATGTGAATGACA  
ATGCCCCGCGGTTCTTCCCCAGCCACTGTGCTGTGGCTGTCTTCGACAACACCACAGT  
GAAGACCCCTGTGGCTGTAGTATTTGCCCGGGATCCCCGACCAAGGCGCCAATGCCCAG  
GTGGTTTACTCTCTGCCGATTACAGCCGAAGGCCACTTTTCCATCGACGCCACCACGG  
GGGTGATCCGCTGGAAAAGCCGCTGCAGGTGAGGCCCCAGGCACCACTGGAGCTCAC  
GGTCCGTGCCCTCTGACCTGGGCACCCCAATACCGCTGTCCACGCTGGGCACCGTCACA  
GTCTCGGTGGTGGGCTAGAAGACTACCTGCCCGTGTTCCTGAACACCGAGCACAGCG  
TGCAGGTGCCCGAGGACGCCCCACCTGGCACGGAGGTGCTGCAGCTGGCCACCCTCAC  
TCGCCCCGGGCGCAGAGAAGACCGGCTACCGCGTGGTCAGCGGGAACGAGCAAGGCAGG  
TTCCGCTTGATGCTCGCACAGGGATCCTGTATGTCAACGCAAGCCTGGACTTTGAGA  
CAAGCCCCAAGTACTTCCTGTCCATTGAGTGCAGCCGGAAGAGCTCCTCTTCCCTCAG  
TGACGTGACCACAGTCATGGTCAACATCACTGATGTCAATGAACACCGGCCCCCAATTC  
CCCCAAGATCCATATAGCACAAAGGTCTTAGAGAATGCCCTTGTGGGTGACGTATCC  
TCACGGTATCAGCGACTGATGAAGATGGACCCCTAAATAGTGACATTACCTATAGCCT  
CATAGAGGGGAACCAGCTTGGGCACTTCACCATTACCCCCAAAAGGGGAGCTACAG  
GTGGCCAAGGCCCTGGACCGGGAACAGGCCTCTAGTTATTTCCCTGAAGCTCCGAGCCA  
CAGACAGTGGGCAGCCTCCACTGCATGAGGACACAGACATCGCTATCCAAGTGGCTGA  
TGTCAATGATAACCCACCGAGATTCTTCCAGCTCAACTACAGCACCCTGTCCAGGAG  
AACTCCCCCATTTGGCAGCAAAGTCCTGCAGCTGATCCTGAGTGACCCAGATTCTCCAG  
AGAATGGCCCCCCCCTACTCGTTTCGAATCACCAAGGGGAACAACGGCTCTGCCTTCCG  
AGTGACCCCGGATGGATGGCTGGTGAAGTGTGAGGGCCTAAGCAGGAGGGCTCAGGAA  
TGGTATCAGCTTCAGATCCAGGCGTCAGACAGTGGCATCCCTCCCCCTCTCGTCTTTGA  
CGTCTGTCCGTGTCCATGTTCACAGAGCAGAGCCACTATGCACCTTCTGTCTCCCACT  
GGAGATCTTCATCACTGTGTTGGAGAGGATGAGTTCCAGGGTGGCATGTGGTGGTAAAGATC  
CATGCCACAGACCGAGACCCCCAGGACACGCTGACCTATAGCCTGGCAGAAAGAGGAGA  
CCCTGGGCAGGCAC'TTCTCAGTGGGTGCGCCTGATGGCAAGATTATCGCCGCCACAGG  
CCTGCCTCGTGGCCACTACTCGTTCAACGTACGGTCAGCGATGGGACCTTCACCACG  
ACTGCTGGGTCCATGTGTACGTGTGGCATGTGGGGCAGGAGGCTCTGCAGCAGGCCA  
TGTGGATGGGCTTCTACCAGCTCACCCCCGAGGAGCTGGTGAGTGACCACTGGCGGAA  
CCTGCAGAGGTTCCCTCAGCCATAAGCTGGACATCAAACGGGCTAACATTCACTTGGCC  
AGCCTCCAGCCTGCAGAGGCCGTGGCTGGTGTGGATGTGCTCCTGGTCTTTGAGGGGC  
ATTCTGGAACCTTCTACGAGTTTCAGGAGCTAGCATCCATCATCACTCACTCAGCCAA  
GGAGATGGAGCATTCACTGGGGGTTTCAGATGCGGTGAGCTATGCCCATGGTGGCTGC  
CAGGGGCCAACCTGCCAGGGTCAAATCTGCCATAACACAGTGCATCTGGACCCCAAGG  
TTGGGGCCACGTACAGCACCAGGCCAGGNTTAACATCCCTAACCCCGCGGCACCACCT  
GCAGAGGAGCTGCTCCTGCAATGGTACTGCTACAAGGTTTCAGTGGTCAGAGCTATGTG  
CGGTACAGGGTCCCAGCGGCTCGGAAGTGGCACATCCATTTCTATCTGAAAACACTCC  
AGCCACAGGCCATTCTTCTATTACCAATGAAACAGCGTCCGTCTCCCTGAAGGGCTT  
TGAAGGCTGCCTGGATGCTGTCTGTGGTCAACGAAGAGGCTCTAGATCTGCTGGCCCT  
GGCAAGACGGTGGCAGGCTTGCTGGAGACACAAGCCCTCACCCAGTGCTGCCTCCACA  
GTGACTACTGCAGCCAGAACACATGCCTCAATGGTGGGAAGTGCTCATGGACCCACGG  
GGCAGGCTATGTCTGCAAATGTCCCCACAGTTCTCTGGGAAGCACTGTGAACAAGGA

	<p>AGGGGAGAACTGTACTTTTGCACCCTGCCTGGAAGGTGGAACCTGTCATCCTCTCCCCCA  AAGGAGCTTCTGTAACTGCCCTCATCCTTACACAGGAGACAGGTGTGAAATGGAGGC  GAGGGGTTGTTTCTGAGGACACTGCCTAGTCACTCCCGAGATCCAAAGGGGGGACTGG  GGGCAGCAGGAGTTACTGATCATCACAGTGGCCGTGGCGTTTATTATCATAAGCACTG  TCGGGCTTCTTCTTCTACTGCCGCCGTGCAAGTCTCACAAGCCTGTGGCCATGGAGGA  CCCAGACCTCCTGGCCAGGAGTGTGGTGTGACACCCAAGCCATGCCCTGCCATCGAG  CTCAACCCATTGAGTGCCAGCTCCTGCAACAACCTCAACCAACTGGAACCCAGCAAGG  CCTCTGTTCCAAATGAACTCGTCACATTTGGACCCAATTCTAAGCAACGGCCAGTGGT  CTGCAGTGTGCCCCCAGACTCCCGCCAGCTGCGGTCCCTTCCCACTCTGACAATGGG  CCTGTCTATTAAGAGAACCTGGTCCAGTGAGGAGATGGTGTACCCTGGCGGAGCCATGG  TCTGGCCCCCTACTTACTCCAGGAACGAACGCTGGGAATACCCCCACTCCGAAGTGAC  TCAGGGCCCTCTGCCGCCCTCGGCTCACCGCCACTCAACCCAGTCTGTATGCCAGAG  CCTAATGGCCTCTATGGGGGCTTCCCTTCCCTTGGAGATGGAAACAAGCGGGGCAC  CTCTCCCACCCCGTTACAGCAACCAGAACCTGGAAGATCTGATGCCCTCTCGGCCCCC  TAGTCCCCGGGAGCGCCTGGTTGCCCCCTGTCTCAATGAGTACACGGCCATCAGCTAC  TACCACTCGCAGTTCCGGCAGGGAGGGGGAGGGCCCTGCCTGGCAGACGGGGGCTACA  AGGGGGTGGGTATGCGCCTCAGCCGAGCTGGGCCCTCTTATGTCTGTGTGAGGTTGA  GGGGGCACCTCTTGACAGGCCAGGGCCAGCCCCGGGTGCCCCCAACTATGAGGGCTCT  GACATGGTGGAGAGTGATTATGGCAGCTGTGAGGAGGTGATGTTCTAGCTTCCCATTCT  CCAGAGCAAGGCAGGCGGGAGGCCAAGGACTGGACTTGGCTTATTTCTTCTGTCTCG  TAGGGGGTGGTGTGAGTGTGGCTGGGAGAGTGGGAGGGAAGCCCTCAGCCAGGCTGT  TGTCCCTTGAAATGTGCTCTTCCAATCCCCACCTAGTCCCTGAGGGTGGAGGGAAGC  TGAGGATAGAGCTCCAGAAACAGCACTAGGGTCCCAGGAGAGGGGCATTCTAGAGCA  GTGACCCTGGAAAACCAGGAACAATTGACTCCCGGGGTGGGCGAGAGACAGGAGGGCT  CCCTGATCTGCCGGCTCTCAGTCCCCGGGGCAGAGCCTGATTGACTGTGCTGGCTCAA  CTTCACCAAGATGCATTCTCATACCTGCCACAGCTCCATTCTGAGGAGGAGGTT  GGTGCCTGACAGACAACCATTACGCGGGCCGTACAGAGGAGCTCTAGAGGGCTGCGTG  GCATCCTCCTAGGGGCTGAGAGGTGAGCAGCAGGGGAGCGGGCAGTCCCCCTCTGCC  CCTGCCCTCAGTCGAGCACTCACTGTGTCTTTGTCAAGTGTCTGCTCCACGTCAGGCAC  TGTGCTTTGCACCGGGGAGAAAATGGTGTATGGAGGGCAACAAGGACTCCGAGGAGCAC  CACCAGGCCTCGGGCCCCAGAGGTCCCCTCCTCAGCCTACACGAGAGGAACGGGGCC  CACCTCAGAGTCACACCACTGGCTGCCAGTCAGGGCCTGCCAGGAGTCTACACAGCTC  TGAACCTTCTTTGTTAAAGAATTGAGACCTCATGGAACCTCTGGGTCTTTCATCCCAAG  TTTCCCAGGCACTTTTGGCCAAAGGAAGGAAGGAACCTAATTCTTCAATTTAAAAATTC  TTAGGCACTTTTGTACCTTGCTGTCTGGATGAGTTTCTCAATGGGATTTTCTTCCC  TAGACACAAGGAAGTCTGAACCTCTATTTAGGGCCGGTTGGAAGCAGGGAGCTGGACC  GCAGTGTCCAGGCTGGACACCTGCCATTGCCTCCTCTCCATTGCAGACGCCTGCCCAT  CAAGTATTACTGCGGCGACTCAACCTATGCATGGAGGGTCAATGTGGGCACATGTCT  ACACATGTGGGTGCCCATGGATAGTACGTGTGTACACATGTGTAGAGTGTATGTAGCC  AGGAGTGGTGGGGACAGAAAGCCTCTGTGGCCTTTGGTGACCTCACCCTCCCTCCCA  CCCAGTCCCTCCCTCTGGTCCACTGCCTTTTTCATATGTGTTGTTTCTGGAGACAGAAG  TCAAAAGGAAGAGCAGTGGAGCCTTGCCACAGGGCTGCTGCTTCATGCGAGAGGGAG  ATGTGTGGGCGAGAGCCAATTTGTGTGAGTGGTTTGTGGCTGTGTGTGTGACTGTGAG  TGTGAGTGACAGATACATAGTTTCATTGGTCAATTTTTTTTAAACAATAAAGTATCTT  TTTTTACTGTT</p>
	<div>ORF Start: at 2</div> <div>ORF Stop: at 12794</div>
	<div>SEQ ID NO: 114</div> <div>4264 aa</div> <div>MW at 469871.7 Da</div>
NOV28b, CG51923-03 Protein Sequence	<p>MTIALLGFAIFLLHCATCEKPLEGILSSSAWHFTSHYNATIYENSSPKTYVESFEKM  GIYLAEPQWAVRYRIISGDVANVFKTEEYVVGNEFCFLRIRTKSSNTALLNREVRDSYT  LI IQATEKTLELEALTRVVHILDQNDLKPLFSPPSYRVTISEDMPKSPICKVTATD  ADLGQNAEFYYAFNTRSEMFAIHPTSGVVTVAGKLNVTWRGKHELQVLAVDRMRKISE  GNGFGSLAALVVHVEPALRKPPAIASVVVTPPDSNDGTTYATVLVDANSSGAEEVESVE  VVGDPGKHFKAIKSYARSNEFSLVSVKIDINWMEYLHGFNLSLQARSGSPYFYQSIR  GFHLPPSKLSSLKFKAIVYRVQLSEFSPPGSRVVMVRVTPAFPNLQYVLKPSSENVGF  KLNARTGLITTTKLMDFHDRHYQLHIRTSPGQASTVVVIDIVDCNNHAPLFNRSSYD  GTLDENIPPGTSVLAVTATDRDHGENGYVTYSIAGPKALPFSIDPYLGIISTSKPMDY  ELMKRIYTFRVRASDWGSPFRREKEVSI FLQLRNLNDNQPMFEEVNCTGSIRQDWPVG  KSIMTMSAIDVDELQNLKYEIVSGNELEYFDLNFHSGVISLKRPFINLTAGQPTSYSLSL  KITASDGKNYASPTTLNITVVKDPHFVPTCDKTGVLTQFTKILHFI GLQNQESSD</p>

EFTSLSTYQINHYTPQFEDHFPQS IDVLESVPINTPLARLAATDPDAGFNGKLVYVI  
 ADGNEEGCFDIELETGLLTVAAPLDYEATNFYILNVTVDLGTQKSSWKLLTVNVKD  
 WNDNAPRFPFPGGYQLTISEDTEVGTTIAELTTKDADSEDNGRVRYTLLSPTEKFSLHP  
 LTGELVVTGHLDRSEPRYILKVEARDQPSKQHLFSVTDLIITLEDVNDNSPQCITE  
 HNRLKVPEDLPPGTVLTFLDASDPDLGPAGEVRYVLMGAGHTFRVDLMTGALILERE  
 LDFERRAGYNLSLWASDGGRLARRTLCHVEVIVLDVNENLHPPHFAS FVHQGVQEN  
 SPSGTQVIVVAAQDDDSGLDGELQYFLRAGTGAAFSINQDTGMIQTLAPLDREFASY  
 YWLTVLAVDRGSVPLSSVTEVYIEVTDANDNPPQMSQAVFYPSIQEDAPVGTSVLQLD  
 AWDPDSSSKGKLTFNITSGNYMGFFMIHPVTGLLSTAQQLDRENKDEHILEVTVDLNG  
 EPSLKSTSRVVVGILDVNDNPPIFSHKLFNVRLPERLSPVSPGPVYRLVASDLDEGLN  
 GRVTYSIEDSYEEAFSIDLVTGVVSSNSTFTAGEYNIILTIKATDSGQPPLSASVRLHI  
 EWIPWPRPSSIPLAFDETYYSFTVMETDPVNHMVGVISVEGRPGFLWFNISGGDKDMD  
 FDIEKTTGSIVIARPLDTRRRSNYNLTVEVTDGSRITATQVHIFMIANINHHRPQFLE  
 TRYEVVPQDTPVPGVELLRVQAIDQDKGKSLIYTIHGSQDPGSASLFQLDPSSGVLVT  
 VGKLDLGSGPSQHTLTVMVRDQEIPIKRNFVWVTIHVEDGNLHPPRFTQLHYEASVPD  
 TIAPGTELLQVRAMDADRGVNAEVHYSLLKGNSEGGFFNINALLGIITLAQKLDQANHA  
 PHTLTVKAEDQGSQWHDLATVIIHVYPSDRSAPIFSKSEYFVEIPESIPVGSPIILLV  
 SAMPSEVITYELREGNKDGVFSMNSYSGLISTQKKLDHEKISSYQKIRGSMAGAF  
 DMVVDI IDENDNAPMFLKSTFVGQISEAAPLYSMIMDKNNNPFVHASDSDKEANS  
 LLVYKILEPEALKFFKIDPSMGTLTIVSEMDYESMPSFQFCVYVHDQGSPLFAPRPA  
 QVIIHVRDVNDSPPRFSEQIYEVAIVGPIHPGMELLMVRASDEDSEVNYSIKTGNAD  
 AVTIHPVTGSISVLNPAFLGLSRKLTIRASDGLYQDTALVKISLTQVLDKSLQFDQDV  
 YWAAVKENLQDRKALVILGAQGNHLDNTLSYFLLNGTDMFHMVQSAGVLQTRGVAFDR  
 EQQDTHELAVEVRDNRTPQORVAQGLVRVSIEDVNDNPPKFKHLPYYTIIQDGTPEGDV  
 LFQVSATDEDLGTNGAVTYEFAEDYTYFRIDPYLGDISLKKPFQALNKYHLKVIAR  
 DGGTPSLQSEEEVLVTVRNKSNPLFQSPYKVRVPENITLYTPILHTQARSPEGLRLI  
 YNIVEEPLMLFTTDFKTGVLTVTGPLDYESKTKHVFTVRATDTALGSFSEATVEVLV  
 EDVNDNPPPTFSQLVYTTISISEGLPAQTPVIQLLASDQDSGRNRDVSQIIVEDGSDVSK  
 FFQINGSTGEMSTVQELDYEAQQHFHVKVRAMDKGDPPLTGETLVVVNVSDINDNPPE  
 FRQPQYEANVSELATCGHLVLKVQAIDPDSRDTSRLEYLILSGNQDRHFFINSSSGII  
 SMFNLCKKHLDSYNLRVGASDGVFRATVPVYINTTNANKYSPEFQQHLYEAELAENA  
 MVGKVIDLLAIDKDSGPYGTIDYTIINKLASEKFSINPNGQIATLQKLDRENSTERV  
 IAIKVMARDGGGRVAFCTVKIILTDENDNPPQFKASEYTVSIQSNVSKDSPVIQVLAY  
 DADEGQNADVITYSVNPEDLVKDVIEINPVTGVVKVKS LVGLENQTLDFFIKAQDGGP  
 PHWNSLVVRLQVVPKKVSLPKFSEPLYTFSAPEDLPEGSEIGIVKAAQAQDPVIYSL  
 VRGTTPESNKDGVSFLDPDTGVIKVRKPMDEHSTKLYQIDVMAHCLQNTDVSLSVSN  
 IQVGDVNDNRPVFEADPYKAVLTENMPVGTSVIQVTAIDKDTGRDGQVS YRLSADPGS  
 NVHELFAIDSESGWITTLQELDCETCQTYHFHVAYDHGQTIQLSSQALVQVSITDEN  
 DNAPRFASEEYRGSVVENSEPGELVATLKTLDADISEQNRQVTCYITEGDPGQFGIS  
 QVGDEWRISSRKTLDRHTAKYLLRV TASDGKFQASVTVEIFVLVDVNDNSPQCSQLLY  
 TGKVHEDVFPGHFILKVSATDLDTNAQITYSLHGPAGAHEFKLDPHTGELTTLTALD  
 RERKDVFNLVAKATDGGGRSCQADITLHVEDVNDNAPRFPFPHCAVAVFDNTTVKTPV  
 AVVFARDPDQGANAVVYSLPDSAEGHFSIDATTGVRIRLEKPLQVRPQAPLELTVRAS  
 DLGTPIPLSTLGTVTVS VVGLEDYLPVFLNTEHSVQVPEDAPPGTEVLQLATLTRPGA  
 EKTGYRVVSGNEQGRFRLDARTGILYVNASLDFETSPKYFLSIECSRKSSSSLSDVTT  
 VMVNITDVNEHRPQFPQDPYSTRVLENALVGDVILTVSATDEDDGPLNSDITYSLIGGN  
 QLGHFTIHPKKGELQVAKALDREQASSYSLKLRATDSGQPPLHEDTDIAIQVADVNDN  
 PPRFFQLNYSTTVQENSPIGSKVLQLILSDPDSPENGPPYSFRITKGNNGSAFRVTPD  
 GWLVTAEGLSRRAQEWYQLQIQASDSGIPPLSSLTSVRVHVTEQSHYAPSALPLEIFI  
 TVGEDEFQGGMVGKI HATDRDPQDTLTYS LAEEETLGRHFSV GADPGKIIAAQGLPRG  
 HYSFNVTVSDGTFTTTAGVHVYVWHVGQEALQQAMWGMGFYQLTPEELVSDHWRNLQRF  
 LSHKLDIKRANIHLASLQPAEAVAGVDVLLVFEGHSGTFYEFQELASIIITHSAKEMEH  
 SVGVQMRSAAMPVPCQGPCTCQGQICHNTVHLDPKVGPITYSTGFQALTS LTPRHHLQRSC  
 SCNGTATRFSGQSYVRYRVPAAARNWHIHFY LKTLQPQAILLFTNETASVSLKGFEGCL  
 DAVVVNEEALDLLAPGKT VAGLLETQALTQCCLHSDYCSQNTCLNGGKCSWTHGAGYV  
 CKCPPQFSGKHCEQGRENCTFAPCLEGGTCILSPKGASCNCPHYTGDRCMEARGCS  
 EGHCLVTPEIQRGDWGQQELLIITVAVAFIIISTVGLLFYCRRCKSHKPVAMEDPDL  
 ARSVGVDTQAMP AIELNPLSASSCNNLNQLEPSKASVPNELVTFGPNSKQRPVVCVSP  
 PRLPPAAVPSHSDNGPVIKRTWSSEEMVYPGGAMVWPPTYSRNERWEYPHSEVTQGPL  
 PPSAHRHSTPVVMPEPNGLYGGFFPLEMENKRAPLPPRYSNQNL EDLMPSPRPPSPRE  
 RLVAPCLNEYTAISYYSQFRQGGGGPCLADGGYKGVGMRLSRAGPSYAVCEVEGAPL

	AGQGQPRVPPNYEGSDMVESDYGSCEEVMF	
	SEQ ID NO: 115	3678 bp
NOV28c, 207756525 DNA Sequence	<p> <b>AAGCTTTATAAGGCTGTCCTCACTGAGAATATGCCAGTGGGGACCTCAGTCATTCAAG</b>  <b>TGACTGCCATTGACAAGGACACTGGGAGAGATGGCCAGGTGAGCTACAGGCTGTCTGC</b>  <b>AGACCCTGGTAGCAATGTCCATGAGCTTTTTGCCATTGACAGTGAGAGTGGTTGGATC</b>  <b>ACCACACTCCAGGAACCTGACTGTGAGACCTGCCAGACTTATCATTTTCATGTGGTGG</b>  <b>CCTATGACCACGGACAGACCATCCAGCTATCCTCTCAGGCCCTGGTTCAGGTCTCCAT</b>  <b>TACAGATGAGAATGACAATGCTCCCCGATTGTCTTCTGAAGAGTACAGAGGATCTGTG</b>  <b>GTTGAGAACAGTGAGCCTGGCGAACTGGTGGCGACTCTAAAGACCCTGGATGCTGACA</b>  <b>TTTCTGAGCAGAACAGGCAGGTACCTGCTACATCACAGAGGGAGACCCCTGGGCCA</b>  <b>GTTTGGCATCAGCCAAGTTGGAGATGAGTGGAGGATTTCTCAAGGAAGACCCTGGAC</b>  <b>CGCGAGCATAACAGCAAGTACTTGCTCAGAGTCACAGCATCTGATGGCAAGTTCAGG</b>  <b>CTTCGGTCACTGTGGAGATCTTTGTCTGGACGTCAATGATAACAGCCCACAGTGTTC</b>  <b>ACAGCTTCTCTATACTGGCAAGGTTTATGAAGATGTATTTCCAGGACACTTCATTTTG</b>  <b>AAGGCTTCTGCCACAGACTTGACACTGATACCAATGCTCAGATCACATATTTCTCTGC</b>  <b>ATGGCCCTGGGGCGCATGAATTCAGCTGGATCCTCATACAGGGGAGCTGACCACACT</b>  <b>CACAGCCCTAGACCGAGAAAGGAAGGATGTGTTCAACCTTGTGGCAAGGCAGCGAT</b>  <b>GGAGGTGGCCGATCGTGCCAGGCAGACATCACCTCCATGTGGAGGATGTGAATGACA</b>  <b>ATGCCCCGCGGTTCTTCCCCAGCCACTGTGCTGTGGCTGTCTTCGACAAACCCACAGT</b>  <b>GAAGACCCCTGTGGCTGTAGTATTTGCCCGGGATCCCGACCAAGGCGCCAATGCCAG</b>  <b>GTGGTTTACTCTCTGCCGATTACGCCGAAGGCCACTTTTCCATCGACGCCACCACGG</b>  <b>GGGTGATCCGCTGGAAAAGCCGCTGCAGGTCAGGCCCCAGGCACCACTGGAGCTCAC</b>  <b>GGTCCGTGCTCTGACCTGGGCACCCCAATACCGCTGTCCACGCTGGGCACCGTCACA</b>  <b>GTCTCGGTGGTGGGCCTAGAAGACTACCTGCCCGTGTTCCTGAACACCGAGCACAGCG</b>  <b>TGCAGGTGCCCCGAGGACGCCCCACCTGGCACGGAGGTGCTGCAGCTGGCCACCCTCAC</b>  <b>TCGCCCCGGCGCAGAGAAGACCGGCTACCGCGTGGTCAGCGGGAACGAGCAAGGCAGG</b>  <b>TTCCGCTGGATGCTCGCACAGGGATCCTGTATGTCAACGCAAGCCTGGACTTTGAGA</b>  <b>CAAGCCCCAAGTACTTCTGTCCATTGAGTGCAGCCGGAAGAGCTCCTCTTCCCTCAG</b>  <b>TGACGTGACCACAGTCATGGTCAACATCACTGATGTCAATGAACACCGGCCCAATTC</b>  <b>CCCCAAGATCCATATAGCACAAGGGTCTTAGAGAATGCCCTTGTGGGTGACGTCATCC</b>  <b>TCACGGTATCAGCGACTGATGAAGATGGACCCCTAAATAGTGACATTACCTATAGCCT</b>  <b>CATAGGAGGGAAACCAGCTTGGGCACTTCACCATTACCCCCAAAAAGGGGGAGCTACAG</b>  <b>GTGGCCAAGGCCCTGGACCGGGAACAGGCCTCTAGTTATTCCTGAAGCTCCGAGCCA</b>  <b>CAGACAGTGGGCAGCCTCCACTGCATGAGGACACAGACATCGCTATCCAAGTGGCTGA</b>  <b>TGTCAATGATAACCCACCGAGATTCTTCAGCTCAACTACAGCACCCTGTCCAGGAG</b>  <b>AACTCCCCCATTGGCAGCAAAGTCTGCAGCTGATCCTGAGTGACCCAGATTCTCCAG</b>  <b>AGAATGGCCCCCTACTCGTTTCGAATACCAAGGGGAACAACGGCTCTGCCTTCCG</b>  <b>AGTGACCCCGGATGGATGGCTGGTGAAGTCTGAGGGCCTAAGTAGGAGGGCTCAGGAA</b>  <b>TGGTATCAGCTTCAGATCCAGGCGTCAGACAGTGGCATCCCTCCCCCTCTCGTCTTCGA</b>  <b>CGTCTGTCCGTGTCCATGTACAGAGCAGAGCCACTATGCACCTTCTGCTCTCCCACT</b>  <b>GGAGATCTTCATCACTGTTGGAGAGGATGAGTTCAGGGTGGCATGGTGGGTAAAGATC</b>  <b>CATGCCACAGACCGAGACCCCCAGGACAGCTGACCTATAGCCTGGCAGAAGAGGAGA</b>  <b>CCCTGGGCAGGCACTTCTCAGTGGGTGCGCCTGATGGCAAGATTATCGCCGCCAGGA</b>  <b>CCTGCCTCGTGGCCACTACTCGTTCAACGTACGGTCAGCGATGGGACCTTACCACG</b>  <b>ACTGCTGGGGTCCATGTGTATGTGTGGCATGTGGGGCAGGAGGCTCTGCAGCAGGCCA</b>  <b>TATGGATGGGCTTCTACCAGCTACCCCCGAGGAGCTGGTGAAGTGAACACTGGCGGAA</b>  <b>CCTGCAGAGGTTCTCAGCCATAAGCTGGACATCAAACGGGCTAACATTCACTTGGCC</b>  <b>AGCCTCCAGCCTGCAGAGGCCGTGGCTGGTGTGGACGTGCTCCTGGTCTTTGAGGGGC</b>  <b>ATTCTGGAACCTTCTACGAGTTTCAGGAGCTAGCATCCATCATCACTCACTCAGCCAA</b>  <b>GGAGATGGAGCATTCACTGGGGGTTTCAAGTGCAGTATGCCCATGGTGGCCCTGC</b>  <b>CAGGGGCCAACCTGCCAGGGTCAAATCTGCCATAACACAGTGCATCTGGACCCCAAGG</b>  <b>TTGGGCCCACGTACAGCACCAGGCTCAGCATCCTAACCCCGCGGACCCCTGCA</b>  <b>GAGGAGCTGCTCCTGCAATGGTACTGCTACAAGGTTCAAGGTCAGTGGTCAAGCTATGTGCGG</b>  <b>TACAGGGCCCCAGCGGCTCGGAACCTGGCACATCCATTTCTATCTGAAAACACTCCAGC</b>  <b>CACAGGCCATTCTTCTATTACCAATGAAACAGCGTCCGTCTCCCTGAAGCTGGCCAG</b>  <b>TGGAGTGGCCAGCTGGAATACCACTGTCTGGGTGGTTCCTATGGAACCTTTCTCTCC</b>  <b>CAGCGCCATGTGAATGACCACGAGTGGCACTCCATCCTGGTGGAGGAGATGGACGCTT</b>  <b>CCATTCGCTGATGGTTGACAGCATGGGCAACACCTCCCTTGTGGTCCCAGAGAACTG</b>  <b>CCGTGGTCTGAGGCCCGAAAGGCACCTCTTGCTGGGCGGCCTCATTTCTGTTGCATTCT</b> </p>	

	TCCTCGAATGTCTCCCAGGGCTTTGAAGGCTGCCTGGATGCTGTCGTGGTCAACGAAG AGGCTCTAGATCTGCTGGCCCCCTGGCAAGACGGTGGCAGGCTTGCTGGAGACACAAGC CCTCACCACAGTGCTGCCTCCACAGTGACTACTGCAGCCAGAACACATGCCTCAATGGT GGGAAGTGCTCATGGACCCACGGGGCAGGCTATGTCTGCAAATGTCCCCACAGTTCT CTGGGAAGCACTGTGAACAAGGAAGGGAGAACTGTACTTTTGCACCCTGCCTGGAAGG TGGAACCTGCATCCTCTCCCCAAAGGAGCTTCTGTAACTGCCCTCATCCTTACACA GGAGACAGGTGTGAAATGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 116	1226 aa	MW at 134061.7 Da
NOV28c, 207756525 Protein Sequence	KLYKAVLTENMPVGTSVIQVTAIDKDTGRDGVSYRLSADPGSNVHELFAIDSESGWI TTLQELDCETCQTYHFHVAYDHGQTIQLSSQALVQVSI TDENDNAPRFASEEYRGSV VENSEPGELVATLKTLDADISEQNROVTCYITEGDPLGQFGISQVGDEWRISSRKTL REHTAKYLLRV TASDGKFQASVTVEIFVLVDVNDNSPQCSQLLYTGKVHEDVFPGHFIL KASATDLDDTDNAQITYSLHGP GAHEFKLDPHTGELTTLTALDRERKDVFNLVAKATD GGGRSCQADITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAQ VVYSLPDSAEGHFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTP I PLSTLGTVT VSVVGLEDYLPVFLNTEHSVQVPEDAPPGTEVLQLATLTRPGA EKTGYRVVSGNEQGR FRLDARTGILYVNASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNITDVNEHRPQF PQDPYSTRVLENALVGDVILTVSATDEDDGPLNSDITYSLIGGNQLGHFTIHPKKGELQ VAKALDREQASSYSLKL RATDSGQPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQE NSPIGSKVLQLILSDPDSPENGGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRR AQE WYQLQIQASDSGIPPLSSSTSVRVHVTEQSHYAPSALPLEIFITVGEDEFQGGMVGKI HATDRDPQD TLTYS LAEEETLGRHFSVGAPDGKI IAAQDLPRGHYSFNVTVSDGTFTT TAGVHVYVWHVGQ EALQQA IWMGFYQLTPEELVSDHWRNLQRFLSHKLDIKRANIHLA SLQPAEAVAGVDVLLVFEGHSGTFYEFQELASII THSAKEMEHSVGVQMR SAMPMVPC QGPTCQGQICHNTVHLDPKVGPTYSTARLSILTPRHHLQRSCSCNGTATRFSGQSYVR YRAPAARNWHIHFY LKTLQPQA ILLFTNETASVSLKLASGVPQLEYHCLGGFYGNLSS QRHVNDHEWHSILVEEMDASIRLMVDSMGNTSLVVPENCRGLRPERHLLGLILLHS SSNVSQGFEGCLDAVVVN EEA LLLAPGKTVAGLLETQALTQCCLHSDYCSQNTCLNG GKCSWTHGAGYVCKCPPQFSGKHCEQGRENCTFAPCLEGGTCILSPKGASCNCPHPYT GDRCEMLE		
	SEQ ID NO: 117	3677 bp	
NOV28d, 207756686 DNA Sequence	GTACCTATAAGGCTGTCTCACTGAGAATATGCCAGTGGGGACCTCAGTCATTCAAGT GACTGCCATTGACAAGGACACTGGGAGAGATGGCCAGGTGAGCTACAGGCTGTCTGCA GACCCTGGTAGCAATGTCCATGAGCTTTTTGCCATTGACAGTGAGAGTGGTTGGATCA CCACACTCCAGGAAC TTGACTGTGAGACCTGCCAGACTTATCATTTTTATGTGGTGGC CTATGACCACGGACAGACCATCCAGCTATCCTCTCAGGCCCTGGTTCAGGTCTCCATT ACAGATGAGAATGACAATGCTCCCCGATTGTCTCTGAAGAGTACAGAGGATCTGTGG TTGAGAACAGTGAGCCTGGCGAACTGGTGGCGACTCTAAAGACCCTGGATGCTGACAT TTCTGAGCAGAACAGGCAGGTACCTGCTACATCAGAGGGAGACCCCTGGGCCAG TTTGGCATCAGCCAAGTTGGAGATGAGTGGAGGATTTCTCAAGGAAGACCCTGGACC GCGAGCATA CAGCCAAGTACTTGCTCAGAGTCACAGCATCTGATGGCAAGTTCCAGGC TTCGGTCACTGTGGAGATCTTTGTCTGACGTCAATGATAACAGCCCACAGTGTTC CAGCTTCTCTATACTGGCAAGGTTTATGAAGATGTATTTCCAGGACACTTCATTTTGA AGGCTTCTGCCACAGACTTGGACACTGATACCAATGCTCAGATCACATATTCTCTGCA TGGCCCTGGGGCGCATGAATTCAAGCTGGATCCTCATACAGGGGAGCTGACCACACTC ACAGCCCTAGACCGAGAAAGGAAGGATGTGTTCAACCTTGTGTGCCAAGGCGACGGATG GAGGTGGCCGATCGTGCCAGGCAGACATCACCCTCCATGTGGAGGATGTGAATGACAA TGCCCCGCGGTTCTTCCCCAGCCACTGTGCTGTGGCTGTCTTCGACAACACCACAGTG AAGACCCTGTGGCTGTAGTATTTGCCCGGGATCCCGACCAAGGCGCCAATGCCAGG TGGTTTACTCTCTGCCGATTGACCCGAAGGCCACTTTTCCATCGACGCCACCACGGG GGTGATCCGCCTGGAAAAGCCGCTGCAGGTCAAGCCCCAGGCACCACTGGAGCTCACG GTCCGTGCCTCTGACCTGGGCACCCCAATACCGCTGTCCACGCTGGGCACCGTCACAG TCTCGGTGGTGGGCCTAGAAGACTACCTGCCCGTGTTCCTGAACACCGAGCACAGCGT GCAGGTGCCCCGAGGACGCCCCACCTGGCACGGAGGTGCTGCAGCTGGCCACCCTCACT CGCCCCGGGCGCAGAGAAGACCGGCTACCGCGTGGTCAGCGGGAACGAGCAAGGCAGGT TCCGCCTGGATGCTCGCACAGGGATCCTGTATGTCAACGCAAGCCTGGACTTTGAGAC AAGCCCCAAGTACTTCTGTCCATTGAGTGCAGCCGGAAGAGCTCCTCTTCCCTCAGT		



	GACGTGACCACAGTCATGGTCAACATCACTGATGTCAATGAACACCGGCCCAATTCC CCCAAGATCCATATAGCACAAAGGGTCTTAGAGAATGCCCTTGTGGGTGACGTCATCCT CACGGTATCAGCGACTGATGAAGATGGACCCCTAAATAGTGACATTACCTATAGCCTC ATAGGAGGGAACCAGCTTGGGCACCTTACCATTACCCCCAAAAGGGGGAGCTACAGG TGGCCAAGGCCCTGGACCGGGAACAGGCCTCTAGTTATTCCCTGAAGCTCCGAGCCAC AGACAGTGGGCAGCCTCCACTGCATGAGGACACAGACATCGCTATCCAAGTGGCTGAT GTCAATGATAACCCACCGAGATTCTTCCAGCTCAACTACAGCACCCTGTCCAGGAGA ACTCCCCCATTGGCAGCAAAGTCTGCAGCTGATCCTGAGTGACCCAGATTCTCCAGA GAATGGCCCCCTTACTCGTTTTCGAATCACCAAGGGGAACAACGGCTCTGCCTTCCGA GTGACCCCGGATGGATGGCTGGTGAAGTCTGAGGGCCTAAGTAGGAGGGCTCAGGAAT GGTATCAGCTTCAGATCCAGGCGTCAGACAGTGGCATCCCTCCCCCTCTCGTCTTCGAC GTCTGTCCGTGTCCATGTACAGAGCAGAGCCACTATGCACCTTCTGCTCTCCCACTG GAGATCTTCATCACTGTTGGAGAGGATGAGTTCCAGGGTGGCATGGTGGGTAAGATCC ATGCCACAGACCGAGACCCCAAGGACACGCTGACCTATAGCCTGGCAGAAGAGGAGAC CCTGGGCAGGCACTTCTCAGTGGGTGCGCCTGATGGCAAGATTATCGCCGCCCAGGAC CTGCCTCGTGGCCACTACTCGTTCAACGTACGGTCAGCGATGGGACCTTACCACGA CTGCTGGGGTCCATGTGTATGTGTGGCATGTGGGGCAGGAGGCTCTGCAGCAGGCCAT ATGGATGGGCTTCTACCAGCTCACCCCGAGGAGCTGGTGAGTGACCACTGGCGGAAC CTGCAGAGGTTCTCAGCCATAAGCTGGACATCAAACGGGCTAACATTCACTTGGCCA GCCTCCAGCCTGCAGAGGCCGTGGCTGGTGTGGACGTGCTCCTGGTCTTTGAGGGGCA TTCTGGAACCTTCTACGAGTTTTCAGGAGCTAGCATCCATCATCACTCACTCAGCCAAG GAGATGGAGCATTCACTGGGGGTTTCAGATGCGGTGAGCTATGCCCATGGTGCCTGCC AGGGGCCAACCTGCCAGGGTCAAATCTGCCATAACACAGTGCATCTGGACCCCAAGGT TGGGCCACGTACAGCACCGCCAGGCTCAGCATCCTAACCCCGCGGCACCACCTGCAG AGGAGCTGCTCCTGCAATGGTACTGCTACAAGGTTCACTGGTGCAGAGCTATGTGCGGT ACAGGGCCCCCAGCGGCTCGAACTGGCACATCCATTCTATCTGAAAACACTCCAGCC ACAGGCCATTCTTCTATTACCAATGAAACAGCGTCCGTCTCCTGTAAGCTGGCCAGT GGAGTGCCCCAGCTGGAATACCCTGTCTGGGTGGTCTTCTATGGAACCTTCTCTCCC AGCGCCATGTGAATGACCACGAGTGGCACTCCATCCTGGTGGAGGAGATGGACGCTTC CATTCGCCTGATGGTTGACAGCATGGGCAACACCTCCCTTGTGGTCCCAGAGAAGTGC CGTGGTCTGAGGCCCGAAAGGCACCTCTTGTCTGGGCGGCTCATTCTGTTGCATTCTT CCTCGAATGTCTCCCAGGGCTTTGAAGGCTGCCTGGATGCTGTCTGGTCAACGAAGA GGCTCTAGATCTGCTGGCCCCCTGGCAAGACGGTGGCAGGCTTGCTGGAGACACAAGCC CTCACCCAGTGTGCTCCACAGTGAAGTACTGCTGAGCCAGAACACATGCCCTCAATGGTG GGAAGTGTCTCATGGACCCACGGGGCAGGCTATGTCTGCAAATGTCCCCACAGTTCTC TGGGAAGCACTGTGAACAAGGAAGGGAGAAGTGTACTTTTGACCCCTGCCTGGAAGGT GGAAGTGTGATCCTCTCCCCCAAAGGAGCTTCTGTAACTGCCCTCATCCTTACACAG GAGACAGGTGTGAAATGCTCGAG		
	ORF Start: at 3	ORF Stop: end of sequence	
	SEQ ID NO: 118	1225 aa	MW at 133921.4 Da
NOV28d, 207756686 Protein Sequence	TYKAVLTENMPVGTSLVIQVTAIDKDTGRDQVSYRLSADPGSNVHELFAIDSESGWIT TLQELDCETCQTYHFHVAYDHGQTIQLSSQALVQVSI TDENDNAPRFASEEYRGSV ENSEPGELVATLKLTLADISEQNRQVTCYITEGDPLGQFGISQVGDEWRISSRKTLDL EHTAKYLLRV TASDGKFQASVTVEIFVLDVNDNSPQCSQLLYTGKVHEDVFPGHFILK ASATDLDTDTNAQITYSLHGPAGHEFKLDPHGTGELTTLTALDRERKDVFNLVAKATDG GGRSCQADITLHVEDVNDNAPRFFPSHCAVAVFDNTTVKTPVAVVFARDPDQGANAV VYSLPDSAEGHFSIDATTGVIRLEKPLQVRPQAPLELTVRASDLGTPILSTLGTVTV SVVGLEDYLPVFLNTEHSVQVPEDAPPGEVLQLATLTRPGAETGYRVVSGNEQGRF RLDARTGILYVNASLDFETSPKYFLSIECSRKSSSSLSDVTTVMVNI TDVNEHRPQFP QDPYSTRVLENALVGDVILTVSATDEDEGPLNSDITYSLIGGNQLGHFTIHPKKGELQV AKALDREQASSYSLKL RATDSGPPLHEDTDIAIQVADVNDNPPRFFQLNYSTTVQEN SPIGSKVLQLILSDPDSPENGPPYSFRITKGNNGSAFRVTPDGWLVTAEGLSRRQEW YQLQIQASDSGIPPLSSSTS VRVHVTEQSHYAPSALPLEIFITVGEDEFQGGMVGKIH ATDRDPQDTLTYS LAEEETLGRHFSVGAPDGKIIAAQDLPRGHYSFNVTVSDGFTTT AGVHVYVWHVGQEAQQAIWMGFYQLTPEELVSDHWRNLQRFLSHKLDIKRANIHLAS LQPAEAVAGVDVLLVFEGHSGTFYEFQELASII THSAKEMEHSVGVQMR SAMPMPVCQ GPTCQGQICHNTVHLDPKVGPTYSTARLSILTPRHHLQRSCSCNGTATRFSGQSYVRY RAPAARNWHIHFYKTLQPAIILFTNETASVSLKLASGVPQLEYHCLGGFYGNLSSQ RNVNDHEWHSILVEEMDASIRLMVDSMGNTSLVVPENCRGLRPERHLLGLILLHSS		



	SNVSQGFEGCLDAVVVNEEALDLLAPGKTVAGLLETQALTQCCLHSDYCSQNTCLNGG KCSWTHGAGYVCKCPPQFSGKHCEQGRENCTFAPCLEGGTCILSPKGASCNCPHPYTG DRCEMLE
--	-------------------------------------------------------------------------------------------------------------------------------------

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

<b>Table 28B. Comparison of NOV28a against NOV28b through NOV28d.</b>		
<b>Protein Sequence</b>	<b>NOV28a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV28b	1..3821 1..3822	3788/3822 (99%) 3790/3822 (99%)
NOV28c	2802..4023 3..1224	1190/1222 (97%) 1191/1222 (97%)
NOV28d	2802..4023 2..1223	1190/1222 (97%) 1191/1222 (97%)

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

<b>Table 28C. Protein Sequence Properties NOV28a</b>	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1030 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 19 and 20

- 5 A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28D.

<b>Table 28D. Geneseq Results for NOV28a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV28a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABG22977	Novel human diagnostic protein #22968 - Homo sapiens, 4591 aa. [WO200175067-A2, 11-OCT-2001]	26..4033 28..4101	1844/4089 (45%) 2647/4089 (64%)	0.0
ABG22977	Novel human diagnostic protein #22968 - Homo sapiens, 4591 aa. [WO200175067-A2, 11-OCT-2001]	26..4033 28..4101	1844/4089 (45%) 2647/4089 (64%)	0.0
AAM52106	Rat fat 3 protein SEQ ID NO 3 - Rattus norvegicus, 4555 aa. [JP2001258573-A, 25-SEP-2001]	10..4169 18..4307	1829/4308 (42%) 2608/4308 (60%)	0.0

ABB71609	Drosophila melanogaster polypeptide SEQ ID NO 41619 - Drosophila melanogaster, 4643 aa. [WO200171042-A2, 27-SEP-2001]	14..3771 41..3869	1303/3874 (33%) 2023/3874 (51%)	0.0
ABG26465	Novel human diagnostic protein #26456 - Homo sapiens, 505 aa. [WO200175067-A2, 11-OCT-2001]	3845..4349 1..505	503/505 (99%) 503/505 (99%)	0.0

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

Table 28E. Public BLASTP Results for NOV28a				
Protein Accession Number	Protein/Organism/Length	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NYQ8	Protocadherin fat 2 - Homo sapiens (Human), 4349 aa.	1..4349 1..4349	4349/4349 (100%) 4349/4349 (100%)	0.0
O88277	MEGF1 - Rattus norvegicus (Rat), 4351 aa.	1..4349 1..4351	3557/4351 (81%) 3915/4351 (89%)	0.0
Q14517	Cadherin-related tumor suppressor homolog precursor (Fat protein homolog) - Homo sapiens (Human), 4590 aa.	26..4033 27..4100	1844/4089 (45%) 2647/4089 (64%)	0.0
Q9WU10	Protocadherin - Rattus norvegicus (Rat), 4589 aa.	3..4033 5..4099	1843/4109 (44%) 2645/4109 (63%)	0.0
Q9QXA3	Mouse fat 1 cadherin - Mus musculus (Mouse), 4587 aa (fragment).	33..4167 35..4315	1890/4317 (43%) 2701/4317 (61%)	0.0

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cadherin	38..139	25/113 (22%) 68/113 (60%)	0.029
cadherin	153..247	28/109 (26%) 69/109 (63%)	5.1e-08
cadherin	367..449	21/107 (20%) 54/107 (50%)	0.55
cadherin	463..553	40/107 (37%) 67/107 (63%)	1.7e-20

cadherin	569..659	26/110 (24%) 65/110 (59%)	2.4e-06
cadherin	720..811	34/107 (32%) 70/107 (65%)	5.3e-24
cadherin	825..916	33/107 (31%) 75/107 (70%)	7.9e-24
cadherin	930..1019	32/107 (30%) 61/107 (57%)	4.2e-12
cadherin	1037..1128	41/107 (38%) 69/107 (64%)	3.2e-21
cadherin	1142..1233	38/107 (36%) 68/107 (64%)	2e-20
cadherin	1247..1337	33/110 (30%) 65/110 (59%)	1.2e-08
cadherin	1354..1438	29/107 (27%) 62/107 (58%)	2.7e-05
cadherin	1453..1546	29/110 (26%) 67/110 (61%)	6.5e-09
cadherin	1560..1651	41/107 (38%) 69/107 (64%)	1.6e-21
cadherin	1665..1749	26/107 (24%) 63/107 (59%)	5.6e-13
cadherin	1763..1861	23/114 (20%) 76/114 (67%)	7.1e-09
cadherin	1877..1959	30/111 (27%) 53/111 (48%)	0.18
cadherin	1973..2061	23/108 (21%) 60/108 (56%)	5.8e-06
cadherin	2075..2164	35/107 (33%) 64/107 (60%)	1.2e-17
cadherin	2176..2263	30/107 (28%) 62/107 (58%)	3.2e-09
cadherin	2277..2370	32/109 (29%) 73/109 (67%)	1.1e-26
cadherin	2384..2472	34/111 (31%) 66/111 (59%)	1.7e-09
cadherin	2486..2576	34/107 (32%) 65/107 (61%)	4.3e-15
cadherin	2590..2682	24/111 (22%) 66/111 (59%)	6.5e-06

cadherin	2696..2786	31/112 (28%) 68/112 (61%)	7.6e-07
cadherin	2802..2897	35/110 (32%) 76/110 (69%)	1.8e-22
cadherin	2911..3002	37/107 (35%) 64/107 (60%)	1.2e-12
cadherin	3016..3104	33/107 (31%) 68/107 (64%)	2.9e-21
cadherin	3119..3209	35/107 (33%) 66/107 (62%)	3e-13
cadherin	3223..3312	31/107 (29%) 68/107 (64%)	7.3e-13
cadherin	3326..3417	43/107 (40%) 70/107 (65%)	5.4e-27
cadherin	3431..3522	35/108 (32%) 74/108 (69%)	3.3e-20
cadherin	3536..3620	25/108 (23%) 56/108 (52%)	0.00082
laminin_G	3800..3924	35/161 (22%) 78/161 (48%)	0.00047
EGF	3951..3983	18/47 (38%) 27/47 (57%)	3.6e-06
EGF	3990..4021	17/47 (36%) 26/47 (55%)	0.00016

### Example B: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate

end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. **PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states

from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

**4. RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue

samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5       **5. Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was  
10 encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species.

These primers were then employed in PCR amplification based on the following  
15 pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, bone marrow, liver, lymphoma.  
20 Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and  
25 with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6.       **Physical Clone:** Exons were predicted by homology and the intron/exon  
30 boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to

further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

#### Molecular Cloning of CG110725-01 (17-290 aa)

The cDNA coding for the mature form of CG110725-01 from residue 17 to 290 of NOV11a was targeted for "in-frame" cloning by PCR. The PCR template is based on the previously identified plasmid.

The oligonucleotide primers in Table BA were used to clone the target cDNA sequence.

**Table BA: Oligonucleotide Primers**

Primers	Sequences	Length	SEQ ID No
F1	5' - GGATCCATACCAGTTAAACAGGCTGATTCTGG - 3'	32	123
R1	5' - CTCGAGATTGACCTCAGAAGATGCACTATCTAATTC - 3'	36	124

For downstream cloning purposes, the forward primer includes an in-frame BamHI restriction site and the reverse primer contains an in-frame Xho I restriction site.

FIS as template: Two parallel PCR reactions were set up using a total of 0.5-1.0 ng human pooled cDNAs as template for each reaction. The pool was composed of 5 micrograms of each of the following human tissue cDNAs: adrenal gland, whole brain, amygdala, cerebellum, thalamus, bone marrow, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, liver, lymphoma, Burkitt's Raji cell line, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small Intestine, spleen, stomach, thyroid, trachea, uterus. When the tissue of expression is known and available, the second PCR was performed using the above primers and 0.5ng-1.0 ng of one of the following human tissue cDNAs: skeleton muscle, testis, mammary gland, adrenal gland, ovary, colon, normal cerebellum, normal adipose, normal skin, bone marrow, brain amygdala, brain hippocampus, brain substantia nigra, brain thalamus, thyroid, fetal lung, fetal liver, fetal brain, kidney, heart, spleen, uterus, pituitary gland, lymph node, salivary gland, small intestine, prostate, placenta, spinal cord, peripheral blood, trachea, stomach, pancreas, hypothalamus.



Two PCR reactions were set up using a total of 1-5 ng of the plasmid that contains the insert for CG110725-01. The reaction mixtures contained 2 microliters of each of the primers (original concentration: 5 pmol/ul), 1 microliter of 10mM dNTP (Clontech Laboratories, Palo Alto CA) and 1 microliter of 50xAdvantage-HF 2 polymerase (Clontech Laboratories) in 50 microliter-reaction volume. The following reaction conditions were used:

PCR condition 1:

- a) 96°C 3 minutes
- b) 96°C 30 seconds denaturation
- c) 60°C 30 seconds, primer annealing
- d) 72°C 6 minutes extension

Repeat steps b-d 15 times

- e) 96°C 15 seconds denaturation
- f) 60°C 30 seconds, primer annealing
- g) 72°C 6 minutes extension

Repeat steps e-g 29 times

- e) 72°C 10 minutes final extension

PCR condition 2:

- a) 96°C 3 minutes
- b) 96°C 15 seconds denaturation
- c) 76°C 30 seconds, primer annealing, reducing the temperature by 1 °C per cycle
- d) 72°C 4 minutes extension

Repeat steps b-d 34 times

- e) 72°C 10 minutes final extension

An amplified product was detected by agarose gel electrophoresis. The fragment was gel-purified and ligated into the pCR2.1 vector (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation. Twelve clones per PCR reaction were picked and sequenced. The inserts were sequenced using vector-specific M13 Forward and M13 Reverse primers and the gene-specific primers shown in Table BB.

**Table BB: Oligonucleotide Primers**

Primers	Sequences	Length	SEQ ID No
1	GATGATGTAGATGACACTGATGATTCT	27	125
2	AAAGCGAGGAGTTGAATGGTGC	22	126

3	ATCTACATCATCAGAGTCGTTTCGAGTCAA	29	127
4	TTTCCATGTGTGAGGTGATGTCCTCGT	27	128

The insert was subsequently cloned into the expression vectors pMel-V5-His and pCEP4-Sec (CuraGen Corporation) by digestion/ligation using the restriction sites BamH I and Xho I.

5 The insert assembly 209934449 was found to encode an open reading frame between residues 17 and 290 of the target sequence of CG110725-01. The cloned insert is 100% identical to the original amino acid sequence. The alignment with CG110725-01 is displayed in a CLUSTAL W (1.7) multiple sequence alignment below. Note that differing amino acids have a white or grey background, and deleted/inserted amino acids can be  
10 detected by a dashed line in the sequence that does not code at that position. The first two and last two amino acids of the insert assembly 209934449 are coded by the primers.

#### Molecular Cloning of CG115187-01:Novel Human Transmembrane Protein

The cDNA coding for a partial ORF of CG115187-01 from residue 247 to 349 of  
15 NOV18a was targeted for "in-frame" cloning by PCR. The PCR template is based on the previously identified plasmid, when available, or on human cDNA(s).

The oligonucleotide primers in Table BC were used to clone the target cDNA sequence.

**Table BC: Oligonucleotide Primers**

Primers	Sequences	Length	SEQ ID No
F3	5' - CACCGGATCC AAGGACATGAA'CCA'CTCTCCAGCACTG - 3'	40	129
R1	5' - GCCCTCGAG GGGACTGTAAGGTGGTGGCTTTTCAAAGG - 3'	39	130

20

For downstream cloning purposes, the forward primer includes an in-frame BamH I restriction site and the reverse primer contains an in-frame Xho I restriction site.

The reaction mixtures contained 2 microliters of each of the primers (original concentration: 5 pmol/ $\mu$ l), 1  $\mu$ l of 10mM dNTP (Clontech Laboratories, Palo Alto CA) and  
25 1  $\mu$ l of Pfu DNA polymerase (Stratagene) in 50 microliter-reaction volume. Conditions used were as described above as PCR condition 1 and PCR condition 2.

An amplified product was detected by agarose gel electrophoresis. The fragment was gel-purified and ligated into the pCR2.1TOPO vector (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation. Twelve clones per PCR reaction were

picked and sequenced. The inserts were sequenced using vector-specific M13 Forward and M13 Reverse primers.

The insert assembly 257788219 was found to encode an open reading frame between residues 247 and 349 of the target sequence of CG115187-01. The cloned insert is 100% identical to the original sequence. The alignment with CG115187-01 is displayed in a ClustalW below. Note that differing amino acids have a white or gray background, and a dashed line indicates deleted or inserted amino acids. The additional amino acids at the ends of the assembly ORF are encoded by the restriction endonuclease sites incorporated into the amplification primers.

## 10 Cloning and Expression of CG110725-04 protein

### Construction of the mammalian expression vector pCEP4/Sec.

The oligonucleotide primers, pSec-V5-His Forward and the pSec-V5-His Reverse were designed to amplify a fragment from the pcDNA3.1-V5His (Invitrogen, Carlsbad, CA) expression vector. The oligonucleotide primers are shown in Table BD. The PCR product was digested with XhoI and ApaI and ligated into the XhoI/ApaI digested pSecTag2 B vector (Invitrogen, Carlsbad CA). The correct structure of the resulting vector, pSecV5His, was verified by DNA sequence analysis. The vector pSecV5His was digested with PmeI and NheI, and the PmeI-NheI fragment was ligated into the BamHI/Klenow and NheI treated vector pCEP4 (Invitrogen, Carlsbad, CA). The resulting vector was named as pCEP4/Sec.

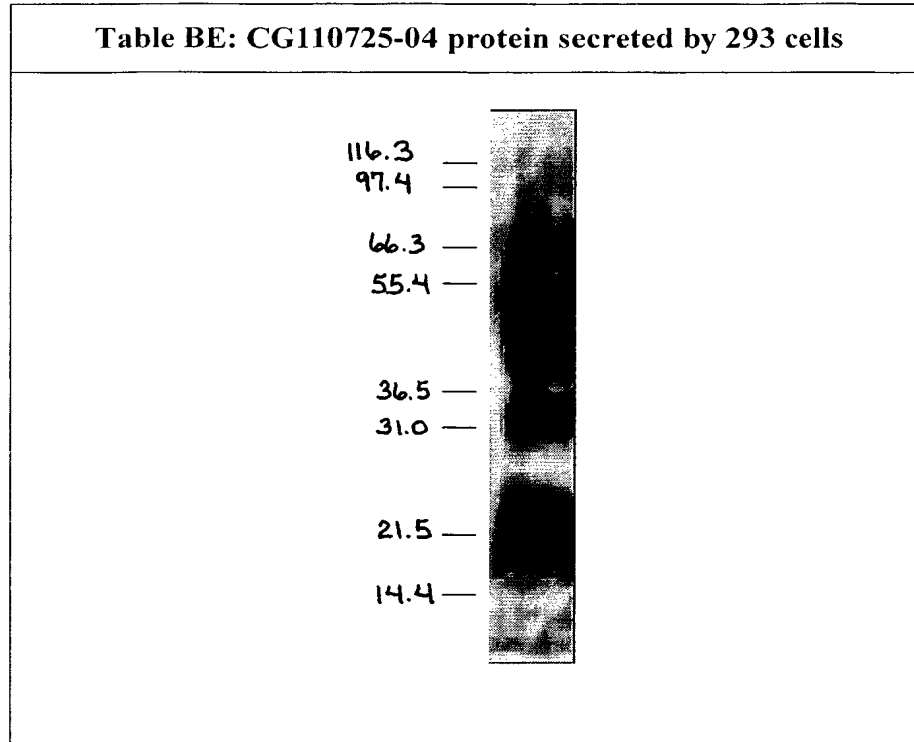
**Table BD: Oligonucleotide Primers**

Primer	Sequence	Length	SEQ ID NO
pSec-V5-His Forward	5' -CTCGTCCTCGAGGGTAAGCCTATCCCTAAC-3'	30	131
pSec-V5-His Reverse	5' -CTCGTCGGGCCCCTGATCAGCGGGTTTAAAC-3'	31	132

### Expression of CG110725-04 in human embryonic kidney 293 cells.

A 0.8 kb BamHI-XhoI fragment containing the CG110725-04 sequence was subcloned into BamHI-XhoI digested pCEP4/Sec to generate plasmid 1323. The resulting plasmid 1323 was transfected into 293 cells using the LipofectaminePlus reagent following the manufacturer's instructions (Gibco/BRL). The cell pellet and supernatant were harvested 72 h post transfection and examined for CG110725-04 expression by Western blot (reducing conditions) using an anti-V5 antibody. Table BE shows that CG110725-04

is expressed as a 35 kDa protein secreted by 293 cells. Some higher molecular weight bands are also visible, which may represent non-reduced form(s) of the protein.



5

### Example C: Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI\_comprehensive\_panel (containing normal tissue and samples from autoimmune/autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS\_neurodegeneration\_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example,  $\beta$ -actin and GAPDH). Normalized RNA (5  $\mu$ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10  $\mu$ g of total RNA were performed in a volume of 20  $\mu$ l and incubated for 60 minutes at 42 °C. This reaction can be scaled up to 50  $\mu$ g of total RNA in a final volume of 100  $\mu$ l. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58 °-60 °C, primer optimal  $T_m$  = 59 °C, maximum primer difference = 2 °C, probe does not have 5'G, probe  $T_m$  must be 10 °C greater than primer  $T_m$ , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by SyntheGen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and

primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No.

4313803) following manufacturer's instructions. Reverse transcription was performed at

48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95 °C for 15 seconds, 60 °C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative

expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95 °C 10 min, then 40 cycles of 95 °C for 15 seconds, 60 °C for 1 minute. Results were analyzed and processed as described previously.

#### **Panels 1, 1.1, 1.2, and 1.3D**

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord,

thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

\* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

#### **General\_screening\_panel\_v1.4, v1.5 and v1.6**

The plates for Panels 1.4, v1.5 and v1.6 include two control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4, v1.5 and v1.6 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, v1.5 and v1.6 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, v1.5 and v1.6 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

**Panels 2D, 2.2, 2.3 and 2.4**

The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include two control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardaïs or Clinomics. The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/ CHTN/Ardaïs/Clinomics). Unmatched RNA samples from tissues without malignancy (normal tissues) were also obtained from Ardaïs or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (*i.e.* immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, *etc.*). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen. General oncology screening panel\_v\_2.4 is an updated version of Panel 2D.

**HASS Panel v 1.0**

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were



obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples. RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

### **Panels 3D and 3.1**

The plates of Panels 3D and 3.1 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature. Oncology\_cell\_line\_screening\_panel\_v3.2 is an updated version of Panel 3. The cell lines in panel 3D, 3.1, 1.3D and oncology\_cell\_line\_screening\_panel\_v3.2 are of the most common cell lines used in the scientific literature.

### **Panels 4D, 4R, and 4.1D**

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial

cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $(5.5 \times 10^{-5}$ M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GM-CSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM

Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml.

Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10µg/ml for 6 and 12-14 hours.

5 CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive  
10 selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and plated at  $10^6$  cells/ml onto Falcon 6 well tissue culture plates that  
15 had been coated overnight with 0.5µg/ml anti-CD28 (Pharmingen) and 3µg/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100µM non essential amino  
20 acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100µM non essential  
25 amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential  
30 amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5µg/ml or anti-CD40 (Pharmingen) at approximately 10µg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 $\mu$ g/ml anti-CD28 (Pharmingen) and 2 $\mu$ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10<sup>5</sup>-10<sup>6</sup> cells/ml in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1 $\mu$ g/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1 $\mu$ g/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 $\mu$ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5x10<sup>5</sup> cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5x10<sup>5</sup> cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1 $\mu$ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), and 10mM Hepes (Gibco). CCD106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were

activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  $10^7$  cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at  $-20^{\circ}\text{C}$  overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 $\mu\text{l}$  of RNase-free water and 35 $\mu\text{l}$  buffer (Promega) 5 $\mu\text{l}$  DTT, 7 $\mu\text{l}$  RNAsin and 8 $\mu\text{l}$  DNase were added. The tube was incubated at  $37^{\circ}\text{C}$  for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at  $-80^{\circ}\text{C}$ .

#### 15 AI\_comprehensive panel\_v1.0

The plates for AI\_comprehensive panel\_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used.

Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebid and two were on phenobarbital.

5 Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1 anti-  
 10 prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI\_comprehensive panel\_v1.0 panel, the following abbreviations are used:

- 15 AI = Autoimmunity
- Syn = Synovial
- Normal = No apparent disease
- Rep22 /Rep20 = individual patients
- RA = Rheumatoid arthritis
- 20 Backus = From Backus Hospital
- OA = Osteoarthritis
- (SS) (BA) (MF) = Individual patients
- Adj = Adjacent tissue
- Match control = adjacent tissues
- 25 -M = Male
- F = Female
- COPD = Chronic obstructive pulmonary disease

#### **Panels 5D and 5I**

30 The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2	Diabetic Hispanic, overweight, not on insulin
Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
Patient 10	Diabetic Hispanic, overweight, on insulin
Patient 11	Nondiabetic African American and overweight
Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U:	Mesenchymal Stem cells, Undifferentiated Adipose
Donor 2 and 3 AM:	Adipose, AdiposeMidway Differentiated
Donor 2 and 3 AD:	Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

5 In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

10 PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

#### **Panel CNSD.01**

15 The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to  
20 confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus,  
25 temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; *e.g.*, Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's  
30 disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.



In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

5 Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

#### **Panel CNS\_Neurodegeneration\_V1.0**

10 The plates for Panel CNS\_Neurodegeneration\_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by  
15 neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided  
20 into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21),  
25 parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and  
30 therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS\_Neurodegeneration\_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; patient not demented but showing severe AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

#### A. CG103191-02 and CG103191-03: chromogranin A

Expression of full length physical clones CG103191-02 and CG103191-03 was assessed using the primer-probe sets Ag6794 and Ag6785, described in Tables AA and AB. Results of the RTQ-PCR runs are shown in Table AC. Please note that Ag6794 is specific to CG103191-02 and Ag6785 is specific to CG103191-03.

**Table AA. Probe Name Ag6794**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID NO
Forward	5' - ggaggctgaggctggaga - 3'	18	762	133
Probe	TET- 5' - ccccgaggaagaaggcccccac - 3' - TAMRA	21	789	134
Reverse	5' - ttcttctcctcggggttcag - 3'	20	817	135

**Table AB. Probe Name Ag6785**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - GAGCCCATGCAGGACAAC - 3'	18	500	136
Probe	FAM- 5' - ACAGTTCCATGAAGCTCTCCTTCCGG - 3' - TAMRA	26	522	137
Reverse	5' - GGCCCTGAAGCCGTA - 3'	16	557	138

**Table AC. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6794, Run 278017436	Tissue Name	Rel. Exp.(%) Ag6794, Run 278017436
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	2.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.5
Squamous cell carcinoma SCC-4	0.6	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	3.6	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0

Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	3.9
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	6.4
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	84.1
Lung ca. NCI-N417	5.4	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	15.4	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	74.7	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	6.0
Lung ca. NCI-H526	2.1	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	1.5
Lung ca. NCI-H460	15.6	Brain (Hippocampus) Pool	2.3
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	14.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	12.2
Liver	0.0	Brain (Thalamus) Pool	6.6
Fetal Liver	0.0	Brain (whole)	12.6
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	100.0
Fetal Kidney	0.0	Pituitary gland Pool	33.9
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	24.7
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6794/Ag6785 Results from one experiment with the CG103191-02 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**General\_screening\_panel\_v1.6 Summary:** Ag6794 Highest expression of the CG103191-02 is detected in adrenal gland (CT=31.6). Moderate to low levels of expression of this gene is also seen in pituitary gland, thyroid and some regions of brain including cerebral cortex, and substantia nigra. Therefore, therapeutic modulation of the protein encoded by this gene may be useful in the treatment of neurological disorders including Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

CG103191-02 codes for a variant of chromogranin A (CgA). CgA is an acidic soluble protein found in the core of secretory vesicles throughout the neuroendocrine system, from which it is coreleased by exocytosis with a variety of amine and peptide hormones and neurotransmitters (O'Connor et al., 1994, Ann N Y Acad Sci 1994 Sep 15;733:36-45, PMID: 7978886). Secretory granules of neuroendocrine cells are inositol 1,4,5-trisphosphate (InsP(3))-sensitive Ca(2+) stores, in which the Ca(2+) storage protein, CGA, couples with InsP(3)-gated Ca(2+) channels (InsP(3)R) located in the granule membrane. The functional aspect of this coupling has been investigated via release studies and planar lipid bilayer experiments in the presence and absence of CGA. CGA drastically increased the release activity of the InsP(3)R by increasing the channel open probability by 9-fold and the mean open time by 12-fold. CGA-coupled InsP(3)Rs are more sensitive to activation than uncoupled receptors. This modulation of InsP(3)R channel activity by CGA appears to be an essential component in the control of intracellular Ca(2+) concentration by secretory granules and may regulate the rate of vesicle fusion and exocytosis (Thrower EC, et al, 2002, J Biol Chem 277:15801-6, PMID: 11842082). A peptide hormone derived from CgA, pancreastatin, is shown to negatively regulate insulin release and exocrine pancreatic secretion (Schmidt WE, Creutzfeldt W, 1991, Acta Oncol 30(4):441-9, PMID: 1854501). Therefore, therapeutic modulation of CgA protein encoded by this gene may be useful in the treatment of neurological and metabolic disorders including diabetes and obesity.

In addition, significant expression of this gene is also seen in a CNS cancer and three lung cancer cell lines. It was shown that CgA is expressed and secreted by a great variety of peptide-producing endocrine neoplasms: pheochromocytoma, parathyroid adenoma, medullary thyroid carcinoma, carcinoids, oat-cell lung cancer, pancreatic islet-cell tumors, and aortic-body tumor (O'Connor and Deftos (1986, New Eng. J. Med.

314:1145-1151, PubMed ID: 3007986). Therefore, expression of the CG103191-02 gene may be used as diagnostic marker to detect the presence of these cancers and also, therapeutic modulation of this gene product may be useful in the treatment of these cancers.

Ag6785 Expression of the CG103191-02 gene is low/undetectable (CTs > 35)

5 across all of the samples on this panel (data not shown).

**Panel 4.1D Summary:** Ag6794/Ag6785 Expression of the CG103191-02 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

#### **B. CG105757-01: Kelch and BTB/POZ domain containing membrane protein**

Expression of gene CG105757-01 was assessed using the primer-probe sets Ag4337  
10 and Ag372, described in Tables BA and BB. Results of the RTQ-PCR runs are shown in Tables BC, BD, BE and BF.

**Table BA. Probe Name Ag4337**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5' - tacaatgctatgtgccaaatcc - 3'	22	325	139
Probe	TET- 5' - catatacacctccgagctggagctca - 3' - TAMRA	26	354	140
Reverse	5' - cagccaccagtgtctcttctgtac - 3'	22	391	141

**Table BB. Probe Name Ag372**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - tgtaacctatgtcacccggcaa - 3'	20	1901	142
Probe	TET- 5' - cgatccagcccgcgttgca - 3' - TAMRA	19	1931	143
Reverse	5' - ctgcgccgtaagtgccattg - 3'	19	1975	144

**Table BC. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4337, Run 224349478	Tissue Name	Rel. Exp.(%) Ag4337, Run 224349478
AD 1 Hippo	16.8	Control (Path) 3 Temporal Ctx	4.4
AD 2 Hippo	27.0	Control (Path) 4 Temporal Ctx	26.2
AD 3 Hippo	5.4	AD 1 Occipital Ctx	11.3
AD 4 Hippo	2.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	96.6	AD 3 Occipital Ctx	4.6
AD 6 Hippo	57.4	AD 4 Occipital Ctx	12.6
Control 2 Hippo	25.0	AD 5 Occipital Ctx	39.2
Control 4 Hippo	10.3	AD 6 Occipital Ctx	11.0
Control (Path) 3 Hippo	6.0	Control 1 Occipital Ctx	3.0
AD 1 Temporal Ctx	13.1	Control 2 Occipital Ctx	85.9

AD 2 Temporal Ctx	28.1	Control 3 Occipital Ctx	13.0
AD 3 Temporal Ctx	4.7	Control 4 Occipital Ctx	3.4
AD 4 Temporal Ctx	18.3	Control (Path) 1 Occipital Ctx	87.1
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	10.7
AD 5 Sup Temporal Ctx	47.3	Control (Path) 3 Occipital Ctx	1.4
AD 6 Inf Temporal Ctx	50.3	Control (Path) 4 Occipital Ctx	10.4
AD 6 Sup Temporal Ctx	50.7	Control 1 Parietal Ctx	5.4
Control 1 Temporal Ctx	4.1	Control 2 Parietal Ctx	44.8
Control 2 Temporal Ctx	50.3	Control 3 Parietal Ctx	23.2
Control 3 Temporal Ctx	11.5	Control (Path) 1 Parietal Ctx	76.8
Control 3 Temporal Ctx	6.0	Control (Path) 2 Parietal Ctx	17.9
Control (Path) 1 Temporal Ctx	52.5	Control (Path) 3 Parietal Ctx	2.7
Control (Path) 2 Temporal Ctx	33.7	Control (Path) 4 Parietal Ctx	39.8

**Table BD. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4337, Run 222550848	Tissue Name	Rel. Exp.(%) Ag4337, Run 222550848
Adipose	6.4	Renal ca. TK-10	14.8
Melanoma* Hs688(A).T	32.3	Bladder	14.6
Melanoma* Hs688(B).T	27.9	Gastric ca. (liver met.) NCI-N87	36.3
Melanoma* M14	44.8	Gastric ca. KATO III	52.1
Melanoma* LOXIMVI	4.6	Colon ca. SW-948	4.2
Melanoma* SK-MEL-5	28.9	Colon ca. SW480	29.3
Squamous cell carcinoma SCC-4	11.0	Colon ca.* (SW480 met) SW620	13.1
Testis Pool	6.5	Colon ca. HT29	6.9
Prostate ca.* (bone met).PC-3	17.0	Colon ca. HCT-116	23.2
Prostate Pool	8.2	Colon ca. CaCo-2	20.7
Placenta	20.6	Colon cancer tissue	12.2
Uterus Pool	3.9	Colon ca. SW1116	6.2
Ovarian ca. OVCAR-3	38.2	Colon ca. Colo-205	2.6
Ovarian ca. SK-OV-3	56.3	Colon ca. SW-48	3.4
Ovarian ca. OVCAR-4	20.9	Colon Pool	21.8
Ovarian ca. OVCAR-5	25.0	Small Intestine Pool	15.1
Ovarian ca. IGROV-1	19.8	Stomach Pool	11.9
Ovarian ca. OVCAR-8	15.3	Bone Marrow Pool	10.0
Ovary	13.7	Fetal Heart	12.3
Breast ca. MCF-7	27.5	Heart Pool	6.9
Breast ca. MDA-MB-231	29.1	Lymph Node Pool	22.7
Breast ca. BT 549	54.7	Fetal Skeletal Muscle	8.0

Breast ca. T47D	54.7	Skeletal Muscle Pool	12.3
Breast ca. MDA-N	17.9	Spleen Pool	8.5
Breast Pool	23.7	Thymus Pool	15.4
Trachea	14.2	CNS cancer (glio/astro) U87-MG	31.2
Lung	5.7	CNS cancer (glio/astro) U-118-MG	100.0
Fetal Lung	23.5	CNS cancer (neuro;met) SK-N-AS	29.1
Lung ca. NCI-N417	3.7	CNS cancer (astro) SF-539	25.0
Lung ca. LX-1	10.0	CNS cancer (astro) SNB-75	57.8
Lung ca. NCI-H146	7.7	CNS cancer (glio) SNB-19	17.8
Lung ca. SHP-77	21.3	CNS cancer (glio) SF-295	73.7
Lung ca. A549	22.4	Brain (Amygdala) Pool	17.2
Lung ca. NCI-H526	2.6	Brain (cerebellum)	52.5
Lung ca. NCI-H23	31.0	Brain (fetal)	55.5
Lung ca. NCI-H460	12.6	Brain (Hippocampus) Pool	16.7
Lung ca. HOP-62	17.4	Cerebral Cortex Pool	22.2
Lung ca. NCI-H522	39.0	Brain (Substantia nigra) Pool	25.0
Liver	3.7	Brain (Thalamus) Pool	29.1
Fetal Liver	14.1	Brain (whole)	29.9
Liver ca. HepG2	8.3	Spinal Cord Pool	9.3
Kidney Pool	32.8	Adrenal Gland	17.4
Fetal Kidney	19.2	Pituitary gland Pool	5.9
Renal ca. 786-0	18.6	Salivary Gland	7.8
Renal ca. A498	5.7	Thyroid (female)	13.1
Renal ca. ACHN	7.7	Pancreatic ca. CAPAN2	7.2
Renal ca. UO-31	12.8	Pancreas Pool	23.5

**Table BE. Panel 1**

Tissue Name	Rel. Exp.(%) Ag372, Run 98747566	Tissue Name	Rel. Exp.(%) Ag372, Run 98747566
Endothelial cells	13.1	Renal ca. 786-0	14.3
Endothelial cells (treated)	6.1	Renal ca. A498	4.4
Pancreas	9.7	Renal ca. RXF 393	14.3
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	15.7
Adrenal gland	27.0	Renal ca. UO-31	8.5
Thyroid	31.0	Renal ca. TK-10	0.4
Salivary gland	9.7	Liver	26.4
Pituitary gland	41.2	Liver (fetal)	11.8
Brain (fetal)	13.0	Liver ca. (hepatoblast) HepG2	25.3
Brain (whole)	11.0	Lung	41.8

Brain (amygdala)	15.1	Lung (fetal)	32.5
Brain (cerebellum)	9.7	Lung ca. (small cell) LX-1	2.6
Brain (hippocampus)	12.9	Lung ca. (small cell) NCI-H69	8.2
Brain (substantia nigra)	10.2	Lung ca. (s.cell var.) SHP-77	2.9
Brain (thalamus)	11.8	Lung ca. (large cell) NCI-H460	22.5
Brain (hypothalamus)	42.9	Lung ca. (non-sm. cell) A549	7.4
Spinal cord	15.7	Lung ca. (non-s.cell) NCI-H23	4.3
glio/astro U87-MG	1.0	Lung ca. (non-s.cell) HOP-62	29.5
glio/astro U-118-MG	12.2	Lung ca. (non-s.cl) NCI-H522	44.8
astrocytoma SW1783	25.9	Lung ca. (squam.) SW 900	10.3
neuro*; met SK-N-AS	66.4	Lung ca. (squam.) NCI-H596	4.7
astrocytoma SF-539	56.3	Mammary gland	41.5
astrocytoma SNB-75	23.3	Breast ca.* (pl.ef) MCF-7	3.5
glioma SNB-19	23.5	Breast ca.* (pl.ef) MDA-MB-231	6.1
glioma U251	18.2	Breast ca.* (pl. ef) T47D	14.6
glioma SF-295	48.3	Breast ca. BT-549	4.0
Heart	32.1	Breast ca. MDA-N	32.8
Skeletal muscle	16.0	Ovary	67.8
Bone marrow	7.0	Ovarian ca. OVCAR-3	13.3
Thymus	11.3	Ovarian ca. OVCAR-4	9.4
Spleen	25.2	Ovarian ca. OVCAR-5	4.0
Lymph node	15.4	Ovarian ca. OVCAR-8	100.0
Colon (ascending)	5.9	Ovarian ca. IGROV-1	21.9
Stomach	10.2	Ovarian ca. (ascites) SK-OV-3	6.3
Small intestine	62.9	Uterus	35.8
Colon ca. SW480	0.0	Placenta	33.9
Colon ca.* SW620 (SW480 met)	2.1	Prostate	31.4
Colon ca. HT29	0.2	Prostate ca.* (bone met) PC-3	23.7
Colon ca. HCT-116	2.4	Testis	16.5
Colon ca. CaCo-2	0.3	Melanoma Hs688(A).T	44.4
Colon ca. HCT-15	5.9	Melanoma* (met) Hs688(B).T	55.9
Colon ca. HCC-2998	12.1	Melanoma UACC-62	70.2
Gastric ca. * (liver met) NCI-N87	6.4	Melanoma M14	45.1
Bladder	51.4	Melanoma LOX IMVI	6.5
Trachea	21.0	Melanoma* (met) SK-MEL-5	27.2
Kidney	32.8	Melanoma SK-MEL-28	0.0
Kidney (fetal)	67.8		

Table BF. Panel 4.1D

Tissue Name	Rel. Exp.(%)	TISSUE NAME	Rel. Exp.(%)
-------------	--------------	-------------	--------------



	Ag4337, Run 184798122		Ag4337, Run 184798122
Secondary Th1 act	36.1	HUVEC IL-1beta	24.0
Secondary Th2 act	75.3	HUVEC IFN gamma	28.1
Secondary Tr1 act	85.3	HUVEC TNF alpha + IFN gamma	13.3
Secondary Th1 rest	22.1	HUVEC TNF alpha + IL4	14.4
Secondary Th2 rest	54.7	HUVEC IL-11	9.5
Secondary Tr1 rest	74.2	Lung Microvascular EC none	56.3
Primary Th1 act	20.7	Lung Microvascular EC TNFalpha + IL-1beta	41.5
Primary Th2 act	81.8	Microvascular Dermal EC none	25.7
Primary Tr1 act	47.6	Microvascular Dermal EC TNFalpha + IL-1beta	19.3
Primary Th1 rest	43.2	Bronchial epithelium TNFalpha + IL1beta	24.8
Primary Th2 rest	35.8	Small airway epithelium none	9.0
Primary Tr1 rest	94.0	Small airway epithelium TNFalpha + IL-1beta	21.5
CD45RA CD4 lymphocyte act	50.3	Coronary artery SMC rest	17.9
CD45RO CD4 lymphocyte act	45.7	Coronary artery SMC TNFalpha + IL-1beta	10.9
CD8 lymphocyte act	39.0	Astrocytes rest	14.1
Secondary CD8 lymphocyte rest	32.1	Astrocytes TNFalpha + IL-1beta	18.4
Secondary CD8 lymphocyte act	17.1	KU-812 (Basophil) rest	5.6
CD4 lymphocyte none	39.5	KU-812 (Basophil) PMA/ionomycin	12.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	81.8	CCD1106 (Keratinocytes) none	30.8
LAK cells rest	46.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	23.2
LAK cells IL-2	36.1	Liver cirrhosis	5.6
LAK cells IL-2+IL-12	19.9	NCI-H292 none	22.1
LAK cells IL-2+IFN gamma	18.3	NCI-H292 IL-4	25.9
LAK cells IL-2+ IL-18	33.7	NCI-H292 IL-9	12.9
LAK cells PMA/ionomycin	9.2	NCI-H292 IL-13	29.9
NK Cells IL-2 rest	74.2	NCI-H292 IFN gamma	16.3
Two Way MLR 3 day	41.8	HPAEC none	18.2
Two Way MLR 5 day	27.4	HPAEC TNF alpha + IL-1 beta	41.8
Two Way MLR 7 day	24.0	Lung fibroblast none	29.3
PBMC rest	15.5	Lung fibroblast TNF alpha + IL-1 beta	19.1
PBMC PWM	19.1	Lung fibroblast IL-4	42.9
PBMC PHA-L	46.0	Lung fibroblast IL-9	50.7

Ramos (B cell) none	1.5	Lung fibroblast IL-13	25.9
Ramos (B cell) ionomycin	2.0	Lung fibroblast IFN gamma	32.8
B lymphocytes PWM	21.6	Dermal fibroblast CCD1070 rest	54.7
B lymphocytes CD40L and IL-4	53.2	Dermal fibroblast CCD1070 TNF alpha	100.0
EOL-1 dbcAMP	45.1	Dermal fibroblast CCD1070 IL-1 beta	27.2
EOL-1 dbcAMP PMA/ionomycin	26.1	Dermal fibroblast IFN gamma	14.4
Dendritic cells none	25.5	Dermal fibroblast IL-4	39.0
Dendritic cells LPS	24.1	Dermal Fibroblasts rest	25.3
Dendritic cells anti-CD40	30.8	Neutrophils TNFa+LPS	2.5
Monocytes rest	28.9	Neutrophils rest	27.0
Monocytes LPS	27.9	Colon	10.0
Macrophages rest	30.4	Lung	11.1
Macrophages LPS	10.9	Thymus	16.8
HUVEC none	18.3	Kidney	33.9
HUVEC starved	30.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4337 This panel confirms the expression of the CG105757-01 gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag4337 Highest expression of this gene is detected in a brain U-118-MG cancer cell line (CT=26.4). High to Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of gastric, colon, lung, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 1 Summary:** Ag4337 Highest expression of this gene is detected in a ovarian cancer OVCAR-8 cell line (CT=26.5). High to Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from liver, gastric, colon, lung, renal, breast, ovarian, melanoma and brain cancers. High to moderate expression is also seen in tissues with metabolic or endocrine functions including pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract and in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Please see panel 1.4 for discussion on utility of this gene.

**Panel 4.1D Summary:** Ag4337 Highest expression of this gene is detected in TNF alpha treated dermal fibroblasts (CT=29). This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

**C. CG108175-01 and CG108175-02 and CG108175-03 and CG108175-04 and CG108175-05: neurexin III-alpha secreted type 1 precursor**

Expression of genes CG108175-01, and variants CG108175-02, CG108175-03, CG108175-04, and CG108175-05 was assessed using the primer-probe sets Ag4351,

Ag6039, Ag6040, Ag6041, and Ag6043 described in Tables CA, CB, CC, CD and CE. Results of the RTQ-PCR runs are shown in Tables CF, CG, CH, CI and CG. Please note CG108175-01 and CG108175-02 are correspond to probe and primer sets Ag4351 and Ag6039 only. In addition, Ag6040 is specific to CG108175-03, Ag6041 is specific to

5 Ag108175-04 and Ag6043 is specific to Ag108175-05.

**Table CA. Probe Name Ag4351**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-gacaaccagtggcacaatgt-3'	20	3605	145
Probe	TET-5'-cgtcatcactcgggacaatagtaaca-3'-TAMRA	26	3625	146
Reverse	5'-taacctgagtgaccacttttgg-3'	22	3671	147

**Table CB. Probe Name Ag6039**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-gacaaccagtggcacaatgt-3'	20	3605	148
Probe	TET-5'-cgtcatcactcgggacaatagtaaca-3'-TAMRA	26	3625	149
Reverse	5'-accacttttgggtgtccactttc-3'	21	3661	150

**Table CC. Probe Name Ag6040**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-caggtaggtcagccagaagc-3'	20	4845	151
Probe	TET-5'-ctagaatcactcgtgccc-3'-TAMRA	22	4875	152
Reverse	5'-agtaaagtgtgaagtgcgcca-3'	24	4908	153

**Table CD. Probe Name Ag6041**

Primers	SEQUENCES	Length	Start Position	SEQ ID No
Forward	5'-GTACAGGTAGGTCAGATAAGAGTCTTTC-3'	28	4842	154
Probe	FAM-5'-CTTCAATCTTCGAAGGTGGCTACAAAGC-3'-TAMRA	28	4872	155
Reverse	5'-TTCGGAGACTTTGTTAGGTCTAAAG-3'	25	4827	156

**Table CE. Probe Name Ag6043**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-CAGGTAGGTCAGTAAGAAATGACAAC-3'	26	4844	157
Probe	FAM-5'-AAAAAGCAAGTTACAAGAATGTGGCAATTCTATTG-3'-TAMRA	36	4872	158
Reverse	5'-ACAAAAGAAAGTTGTGTAAGAATGCT-3'	26	4914	159

**Table CF. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4351, Run 224367441	Rel. Exp.(%) Ag6039, Run 225249579	Rel. Exp.(%) Ag6040, Run 225249580	Rel. Exp.(%) Ag6041, Run 225249580
AD 1 Hippo	27.4	24.8	2.0	3.4
AD 2 Hippo	44.8	14.9	12.3	23.0
AD 3 Hippo	23.7	1.7	0.5	0.8
AD 4 Hippo	17.0	4.3	2.7	2.6
AD 5 Hippo	100.0	78.5	59.9	76.4
AD 6 Hippo	54.3	23.3	24.7	28.4
Control 2 Hippo	37.9	15.1	23.3	27.0
Control 4 Hippo	1.7	2.6	0.3	0.7
Control (Path) 3 Hippo	18.4	1.5	0.2	0.6
AD 1 Temporal Ctx	0.0	9.2	2.5	2.7
AD 2 Temporal Ctx	40.9	21.0	13.1	20.6
AD 3 Temporal Ctx	3.4	1.6	0.5	1.5
AD 4 Temporal Ctx	27.0	13.6	4.5	10.4
AD 5 Inf Temporal Ctx	36.9	82.4	48.3	58.7
AD 5 Sup Temporal Ctx	59.5	25.0	12.6	23.1
AD 6 Inf Temporal Ctx	60.3	20.0	20.7	24.3
AD 6 Sup Temporal Ctx	44.8	24.1	26.8	29.3
Control 1 Temporal Ctx	7.7	1.0	0.5	1.2
Control 2 Temporal Ctx	49.0	31.0	25.5	41.3
Control 3 Temporal Ctx	11.4	9.9	6.8	8.0
Control 3 Temporal Ctx	15.5	3.1	0.9	1.8
Control (Path) 1 Temporal Ctx	51.4	66.0	68.8	86.2
Control (Path) 2 Temporal Ctx	30.8	40.6	30.1	28.8
Control (Path) 3 Temporal Ctx	8.5	1.5	0.3	1.3
Control (Path) 4 Temporal Ctx	45.1	26.1	24.1	20.8
AD 1 Occipital Ctx	28.3	9.4	6.0	8.2
AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0	0.0
AD 3 Occipital Ctx	32.1	1.8	0.2	1.9
AD 4 Occipital Ctx	36.3	11.0	11.7	15.6
AD 5 Occipital Ctx	46.0	32.5	7.0	47.0
AD 6 Occipital Ctx	48.0	15.6	46.3	16.5
Control 1 Occipital Ctx	14.0	0.6	0.2	0.5
Control 2 Occipital Ctx	48.3	51.1	57.4	60.7
Control 3 Occipital Ctx	23.7	15.1	6.6	9.0
Control 4 Occipital Ctx	14.4	2.3	0.4	0.9

Control (Path) 1 Occipital Ctx	55.1	100.0	100.0	100.0
Control (Path) 2 Occipital Ctx	15.0	13.0	4.0	11.5
Control (Path) 3 Occipital Ctx	9.2	1.3	0.2	0.5
Control (Path) 4 Occipital Ctx	25.9	16.4	7.3	9.8
Control 1 Parietal Ctx	13.6	1.9	0.5	1.7
Control 2 Parietal Ctx	47.0	23.7	9.9	13.8
Control 3 Parietal Ctx	26.8	16.6	9.9	16.3
Control (Path) 1 Parietal Ctx	65.5	95.9	71.7	95.4
Control (Path) 2 Parietal Ctx	29.9	19.5	12.5	14.7
Control (Path) 3 Parietal Ctx	7.6	1.6	0.8	0.8
Control (Path) 4 Parietal Ctx	51.1	40.1	49.7	45.9

**Table CG. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4351, Run 222523515	Tissue Name	Rel. Exp.(%) Ag4351, Run 222523515
Adipose	0.3	Renal ca. TK-10	1.7
Melanoma* Hs688(A).T	0.0	Bladder	0.3
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.1
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.1	Colon cancer tissue	0.3
Uterus Pool	0.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.1
Ovary	0.3	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.1
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.4
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.1
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.2

Breast ca. MDA-N	0.0	Spleen Pool	0.1
Breast Pool	0.1	Thymus Pool	2.9
Trachea	0.7	CNS cancer (glio/astro) U87-MG	16.4
Lung	0.1	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	3.8	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	1.1
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	12.2
Lung ca. NCI-H146	6.8	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.1
Lung ca. A549	0.0	Brain (Amygdala) Pool	41.5
Lung ca. NCI-H526	0.2	Brain (cerebellum)	27.5
Lung ca. NCI-H23	0.0	Brain (fetal)	100.0
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	38.7
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	79.0
Lung ca. NCI-H522	3.2	Brain (Substantia nigra) Pool	56.6
Liver	0.0	Brain (Thalamus) Pool	82.4
Fetal Liver	0.6	Brain (whole)	83.5
Liver ca. HepG2	0.0	Spinal Cord Pool	28.3
Kidney Pool	0.1	Adrenal Gland	0.7
Fetal Kidney	2.8	Pituitary gland Pool	2.5
Renal ca. 786-0	2.3	Salivary Gland	1.5
Renal ca. A498	0.0	Thyroid (female)	0.1
Renal ca. ACHN	4.3	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.1

**Table CH. General\_screening\_panel\_v1.5**

Tissue Name	Rel. Exp.(%) Ag6039, Run 228768034	Rel. Exp.(%) Ag6040, Run 228768035	Rel. Exp.(%) Ag6041, Run 228768093
Adipose	0.2	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	0.2
Melanoma* Hs688(B).T	0.2	0.1	1.8
Melanoma* M14	0.0	0.1	0.6
Melanoma* LOXIMVI	0.0	0.0	0.0
Melanoma* SK-MEL-5	0.3	0.1	0.2
Squamous cell carcinoma SCC-4	0.0	0.0	0.0
Testis Pool	1.7	0.1	0.6
Prostate ca.* (bone met) PC-3	0.0	0.0	0.0
Prostate Pool	0.0	0.2	0.9
Placenta	0.0	0.0	0.1

Uterus Pool	0.1	0.1	0.6
Ovarian ca. OVCAR-3	0.0	0.4	0.8
Ovarian ca. SK-OV-3	0.0	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.1	0.5
Ovarian ca. OVCAR-5	0.0	0.0	0.0
Ovarian ca. IGROV-1	0.0	0.0	0.0
Ovarian ca. OVCAR-8	0.0	0.0	0.0
Ovary	0.1	0.0	0.0
Breast ca. MCF-7	0.0	0.3	2.0
Breast ca. MDA-MB-231	0.0	0.0	0.0
Breast ca. BT 549	0.0	0.2	0.8
Breast ca. T47D	0.0	0.0	0.0
Breast ca. MDA-N	0.0	0.0	0.0
Breast Pool	1.2	1.0	5.5
Trachea	0.7	0.2	0.8
Lung	0.3	0.0	0.3
Fetal Lung	3.3	0.5	1.2
Lung ca. NCI-N417	0.1	0.0	0.0
Lung ca. LX-1	0.0	0.0	0.0
Lung ca. NCI-H146	7.6	0.0	0.2
Lung ca. SHP-77	0.0	0.0	0.1
Lung ca. A549	0.0	0.0	0.1
Lung ca. NCI-H526	0.0	0.0	0.0
Lung ca. NCI-H23	0.0	0.0	0.0
Lung ca. NCI-H460	1.0	0.1	0.5
Lung ca. HOP-62	0.2	0.0	0.2
Lung ca. NCI-H522	2.8	0.0	0.1
Liver	0.0	0.0	0.0
Fetal Liver	0.6	0.0	0.1
Liver ca. HepG2	0.0	0.0	0.1
Kidney Pool	0.3	0.4	3.2
Fetal Kidney	2.2	0.8	0.9
Renal ca. 786-0	2.0	1.4	4.4
Renal ca. A498	0.0	0.0	0.3
Renal ca. ACHN	3.7	0.0	0.0
Renal ca. UO-31	0.1	0.3	1.3
Renal ca. TK-10	2.0	0.2	0.6
Bladder	0.4	0.1	0.3
Gastric ca. (liver met.) NCI-N87	0.0	0.0	0.0



Gastric ca. KATO III	0.0	0.0	0.0
Colon ca. SW-948	0.0	0.1	1.2
Colon ca. SW480	0.0	0.1	0.1
Colon ca.* (SW480 met) SW620	0.0	0.0	0.0
Colon ca. HT29	0.0	0.0	0.0
Colon ca. HCT-116	0.0	0.0	0.0
Colon ca. CaCo-2	0.0	0.9	1.4
Colon cancer tissue	0.3	0.0	0.1
Colon ca. SW1116	0.0	0.0	0.0
Colon ca. Colo-205	0.0	0.0	0.0
Colon ca. SW-48	0.0	0.1	0.5
Colon Pool	0.3	1.2	4.4
Small Intestine Pool	0.0	0.3	0.8
Stomach Pool	0.9	0.2	0.7
Bone Marrow Pool	0.2	0.3	0.6
Fetal Heart	0.1	0.1	0.0
Heart Pool	0.0	0.3	1.0
Lymph Node Pool	0.3	0.6	1.8
Fetal Skeletal Muscle	0.2	0.0	0.0
Skeletal Muscle Pool	0.0	0.0	0.1
Spleen Pool	0.1	0.3	1.2
Thymus Pool	2.4	0.2	0.3
CNS cancer (glio/astro) U87-MG	20.2	0.3	0.6
CNS cancer (glio/astro) U-118-MG	0.1	0.0	0.0
CNS cancer (neuro;met) SK-N-AS	0.0	0.0	0.0
CNS cancer (astro) SF-539	0.7	0.0	0.0
CNS cancer (astro) SNB-75	11.7	0.8	1.9
CNS cancer (glio) SNB-19	0.0	0.0	0.1
CNS cancer (glio) SF-295	0.2	0.0	0.1
Brain (Amygdala) Pool	40.9	6.3	6.1
Brain (cerebellum)	65.1	100.0	100.0
Brain (fetal)	100.0	4.6	5.9
Brain (Hippocampus) Pool	42.6	7.3	8.4
Cerebral Cortex Pool	82.9	13.9	15.1
Brain (Substantia nigra) Pool	57.0	8.1	8.3
Brain (Thalamus) Pool	77.9	12.9	13.8
Brain (whole)	84.7	14.3	17.6
Spinal Cord Pool	32.3	2.6	4.1
Adrenal Gland	1.0	0.5	0.8

Pituitary gland Pool	3.5	0.1	0.8
Salivary Gland	1.0	0.0	0.1
Thyroid (female)	0.1	0.0	0.0
Pancreatic ca. CAPAN2	0.0	0.0	0.0
Pancreas Pool	0.4	0.3	0.8

**Table CI. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4351, Run 186363781	Rel. Exp.(%) Ag6040, Run 225158985	Rel. Exp.(%) Ag6041, Run 225159636
Secondary Th1 act	0.0	0.0	0.0
Secondary Th2 act	0.0	0.0	6.2
Secondary Tr1 act	0.0	0.0	0.0
Secondary Th1 rest	0.0	0.0	0.0
Secondary Th2 rest	0.0	0.0	0.0
Secondary Tr1 rest	0.0	0.0	0.0
Primary Th1 act	0.0	0.0	0.0
Primary Th2 act	0.0	0.0	0.0
Primary Tr1 act	0.0	0.0	0.0
Primary Th1 rest	0.0	0.0	0.0
Primary Th2 rest	0.0	0.0	0.0
Primary Tr1 rest	0.0	0.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	0.0
CD8 lymphocyte act	0.0	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	0.0
Secondary CD8 lymphocyte act	0.0	0.0	0.0
CD4 lymphocyte none	0.0	0.0	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	0.0
LAK cells rest	0.7	0.0	0.0
LAK cells IL-2	0.0	0.0	0.0
LAK cells IL-2+IL-12	0.0	0.0	0.0
LAK cells IL-2+IFN gamma	0.0	0.0	0.0
LAK cells IL-2+ IL-18	0.0	0.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	0.0
NK Cells IL-2 rest	0.0	0.0	0.0
Two Way MLR 3 day	0.0	0.0	0.0
Two Way MLR 5 day	0.0	0.0	0.0
Two Way MLR 7 day	0.0	0.0	0.0
PBMC rest	0.0	0.0	0.0

PBMC PWM	0.0	0.0	0.0
PBMC PHA-L	0.9	0.0	0.0
Ramos (B cell) none	0.0	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	0.0
B lymphocytes PWM	0.9	0.0	0.0
B lymphocytes CD40L and IL-4	0.0	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	0.0
Dendritic cells none	0.0	0.0	0.0
Dendritic cells LPS	1.2	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	0.0
Monocytes rest	0.0	0.0	0.0
Monocytes LPS	0.0	0.0	0.0
Macrophages rest	2.0	0.0	4.9
Macrophages LPS	0.0	0.0	0.0
HUVEC none	0.0	7.0	28.3
HUVEC starved	0.0	15.5	31.6
HUVEC IL-1beta	0.0	0.0	23.3
HUVEC IFN gamma	0.0	15.8	39.4
HUVEC TNF alpha + IFN gamma	1.7	8.1	68.4
HUVEC TNF alpha + IL4	0.0	7.2	23.8
HUVEC IL-11	0.0	0.0	16.8
Lung Microvascular EC none	0.0	0.0	0.0
Lung Microvascular EC TNFalpha + IL-1beta	0.8	0.0	0.0
Microvascular Dermal EC none	0.0	0.0	0.0
Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0	0.0
Bronchial epithelium TNFalpha + IL1beta	0.0	0.0	0.0
Small airway epithelium none	0.9	0.0	0.0
Small airway epithelium TNFalpha + IL-1beta	0.0	0.0	3.6
Coronary artery SMC rest	0.0	0.0	7.2
Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0	3.2
Astrocytes rest	0.8	7.2	12.4
Astrocytes TNFalpha + IL-1beta	2.5	5.3	29.5
KU-812 (Basophil) rest	0.0	0.0	0.0
KU-812 (Basophil) PMA/ionomycin	0.0	0.0	0.0
CCD1106 (Keratinocytes) none	0.0	0.0	0.0
CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0	0.0
Liver cirrhosis	0.0	6.0	23.1
NCI-H292 none	0.0	0.0	0.0

NCI-H292 IL-4	0.0	0.0	0.0
NCI-H292 IL-9	0.0	0.0	0.0
NCI-H292 IL-13	0.0	0.0	0.0
NCI-H292 IFN gamma	0.0	0.0	0.0
HPAEC none	0.0	5.6	13.5
HPAEC TNF alpha + IL-1 beta	0.0	0.0	14.1
Lung fibroblast none	0.8	2.2	6.1
Lung fibroblast TNF alpha + IL-1 beta	0.8	0.0	9.9
Lung fibroblast IL-4	0.8	1.4	6.4
Lung fibroblast IL-9	0.0	9.4	36.5
Lung fibroblast IL-13	1.9	2.4	10.1
Lung fibroblast IFN gamma	1.8	0.0	13.1
Dermal fibroblast CCD1070 rest	2.5	0.0	0.0
Dermal fibroblast CCD1070 TNF alpha	1.8	0.0	0.0
Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0	0.0
Dermal fibroblast IFN gamma	6.5	1.2	3.5
Dermal fibroblast IL-4	1.8	0.0	0.0
Dermal Fibroblasts rest	0.0	3.3	0.0
Neutrophils TNFa+LPS	0.0	0.0	0.0
Neutrophils rest	3.2	0.0	0.0
Colon	3.8	5.0	69.1
Lung	18.2	0.0	10.9
Thymus	33.9	14.8	16.2
Kidney	100.0	100.0	100.0
			0.8

**Table CJ. Panel CNS\_1.1**

Tissue Name	Rel. Exp.(%) Ag4351, Run 195308642	Tissue Name	Rel. Exp.(%) Ag4351, Run 195308642
Cing Gyr Depression2	10.6	BA17 PSP2	15.5
Cing Gyr Depression	6.1	BA17 PSP	30.1
Cing Gyr PSP2	3.3	BA17 Huntington's2	10.9
Cing Gyr PSP	13.7	BA17 Huntington's	35.6
Cing Gyr Huntington's2	21.2	BA17 Parkinson's2	75.8
Cing Gyr Huntington's	39.2	BA17 Parkinson's	34.6
Cing Gyr Parkinson's2	26.1	BA17 Alzheimer's2	7.5
Cing Gyr Parkinson's	33.4	BA17 Control2	71.2
Cing Gyr Alzheimer's2	4.9	BA17 Control	70.7

Cing Gyr Alzheimer's	18.9	BA9 Depression2	10.0
Cing Gyr Control2	38.2	BA9 Depression	4.2
Cing Gyr Control	73.7	BA9 PSP2	6.8
Temp Pole Depression2	2.9	BA9 PSP	12.0
Temp Pole PSP2	3.1	BA9 Huntington's2	14.8
Temp Pole PSP	1.3	BA9 Huntington's	37.9
Temp Pole Huntington's	23.2	BA9 Parkinson's2	71.7
Temp Pole Parkinson's2	32.8	BA9 Parkinson's	37.6
Temp Pole Parkinson's	29.3	BA9 Alzheimer's2	11.1
Temp Pole Alzheimer's2	7.7	BA9 Alzheimer's	3.8
Temp Pole Alzheimer's	5.8	BA9 Control2	81.2
Temp Pole Control2	45.7	BA9 Control	20.2
Temp Pole Control	13.5	BA7 Depression	6.9
Glob Palladus Depression	2.7	BA7 PSP2	37.6
Glob Palladus PSP2	3.0	BA7 PSP	42.3
Glob Palladus PSP	3.4	BA7 Huntington's2	48.6
Glob Palladus Parkinson's2	15.9	BA7 Huntington's	45.4
Glob Palladus Parkinson's	75.8	BA7 Parkinson's2	40.3
Glob Palladus Alzheimer's2	4.7	BA7 Parkinson's	15.6
Glob Palladus Alzheimer's	8.2	BA7 Alzheimer's2	7.5
Glob Palladus Control2	10.9	BA7 Control2	31.9
Glob Palladus Control	7.9	BA7 Control	45.1
Sub Nigra Depression2	4.3	BA4 Depression2	10.3
Sub Nigra Depression	7.9	BA4 Depression	11.6
Sub Nigra PSP2	6.7	BA4 PSP2	23.3
Sub Nigra Huntington's2	33.7	BA4 PSP	9.4
Sub Nigra Huntington's	23.7	BA4 Huntington's2	4.9
Sub Nigra Parkinson's2	35.8	BA4 Huntington's	29.7
Sub Nigra Alzheimer's2	7.8	BA4 Parkinson's2	100.0
Sub Nigra Control2	18.9	BA4 Parkinson's	54.7
Sub Nigra Control	15.6	BA4 Alzheimer's2	3.4
BA17 Depression2	22.8	BA4 Control2	52.1
BA17 Depression	10.5	BA4 Control	21.9

**AI\_comprehensive\_Panel\_1.0 Summary:** Ag6043 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4351/Ag6039/Ag6040/Ag6041

Four experiments with three different probe and primer sets produce results in excellent

5 agreement. This gene is not differentially expressed in Alzheimer's disease. However, this

expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

**General\_screening\_panel\_v1.4 Summary:** Ag4351 This gene appears to be preferentially expressed in the brain, with highest expression in a fetal brain sample

5 (CT=28.1). This gene encodes a protein with homology to neurexin, a neuronal cell surface molecule that is involved in synaptogenesis and intercellular signaling. Based on analysis with PFAM, this protein contains both epidermal growth factor-like sequences and domains homologous to the G domain repeats of laminin A supporting a role for this protein in cell-cell interactions. Neurexin has been implicated in synapse formation and may thus  
10 influence learning, memory, and behaviour (Scheiffele P. Cell 2000 Jun 9;101(6):657-69). Neuroxin is also a receptor for the potent neurotoxin alpha-latrotoxin (Geppert M. J Biol Chem 1998 Jan 16;273(3):1705-10). Thus, based on the results seen in this panel, expression of this gene could be used to differentiate between brain tissue and non-neuronal tissue. Furthermore, since this protein is homologous to neurexin, modulation of the  
15 expression or function of this gene may be useful in the treatment of neurodegenerative disorders, and specifically in directing compensatory synaptogenesis in response to neuron death in spinal cord or brain trauma, stroke, Alzheimer's, Parkinson's or Huntington's diseases, or spinocerebellar ataxia.

**General\_screening\_panel\_v1.5 Summary:** Ag6039/Ag6040/Ag6041 Expression  
20 of this gene is highly specific to the brain, in agreement with expression in Panel 1.4. In addition, the experiments with Ag6040 and Ag6041, which are specific to CG108175-03 and CG108175-04 respectively, show significantly higher levels of expression in the cerebellum (CTs=27.7) when compared to expression in other regions of the CNS. This may suggest that this variant is more highly expressed in the cerebellum. Thus, these genes  
25 may also be useful as specific targets of drugs for the treatment of CNS disorders that have this brain region as the site of pathology, such as autism and the ataxias. Ag6043 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel 4.1D Summary:** Ag4351/Ag6040 Expression of this gene is seen at low but  
30 significant levels in normal tissue samples from kidney, thymus, and lung. Thus, expression of this gene could be used to differentiate these samples from other samples on this panel and as a marker of kidney tissue. The expression in normal tissues suggests that this gene product may be involved in maintaining the normal homeostasis of these tissues. Modulation of this gene product may therefore reduce or eliminate symptoms in patients

with autoimmune diseases that affect these organs. Ag6039/Ag6041 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run. Ag6043 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- 5 **Panel CNS\_1.1 Summary:** Ag4351 This expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

#### D. CG108624-01: protocadherin 68 variant

- 10 Expression of gene CG108624-01 was assessed using the primer-probe set Ag4366, described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB, DC, DD and DE.

**Table DA. Probe Name Ag4366**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-tgagcactatctccatcatcct-3'	22	2140	160
Probe	TET-5'-atgatcaccatcgccgtcaagtg-3'-TAMRA	23	2172	161
Reverse	5'-agttgtaagtgcggatctcctt-3'	22	2208	162

**Table DB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4366, Run 224376002	Tissue Name	Rel. Exp.(%) Ag4366, Run 224376002
AD 1 Hippo	24.1	Control (Path) 3 Temporal Ctx	10.0
AD 2 Hippo	58.2	Control (Path) 4 Temporal Ctx	39.2
AD 3 Hippo	12.5	AD 1 Occipital Ctx	20.7
AD 4 Hippo	18.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	57.0	AD 3 Occipital Ctx	9.7
AD 6 Hippo	59.9	AD 4 Occipital Ctx	18.2
Control 2 Hippo	40.1	AD 5 Occipital Ctx	31.0
Control 4 Hippo	16.4	AD 6 Occipital Ctx	18.7
Control (Path) 3 Hippo	9.7	Control 1 Occipital Ctx	4.7
AD 1 Temporal Ctx	22.1	Control 2 Occipital Ctx	31.6
AD 2 Temporal Ctx	38.2	Control 3 Occipital Ctx	20.3
AD 3 Temporal Ctx	9.9	Control 4 Occipital Ctx	8.0
AD 4 Temporal Ctx	29.7	Control (Path) 1 Occipital Ctx	92.7
AD 5 Inf Temporal Ctx	82.4	Control (Path) 2 Occipital Ctx	13.3
AD 5 Sup Temporal Ctx	65.1	Control (Path) 3 Occipital Ctx	5.4
AD 6 Inf Temporal Ctx	56.3	Control (Path) 4 Occipital Ctx	18.2

AD 6 Sup Temporal Ctx	39.5	Control 1 Parietal Ctx	13.2
Control 1 Temporal Ctx	11.5	Control 2 Parietal Ctx	64.6
Control 2 Temporal Ctx	40.3	Control 3 Parietal Ctx	22.7
Control 3 Temporal Ctx	20.0	Control (Path) 1 Parietal Ctx	93.3
Control 3 Temporal Ctx	17.2	Control (Path) 2 Parietal Ctx	28.3
Control (Path) 1 Temporal Ctx	100.0	Control (Path) 3 Parietal Ctx	9.7
Control (Path) 2 Temporal Ctx	48.6	Control (Path) 4 Parietal Ctx	57.0

**Table DC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4366, Run 222543454	Tissue Name	Rel. Exp.(%) Ag4366, Run 222543454
Adipose	3.1	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	12.2
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.4	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	1.1	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	10.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	1.0	Colon ca. CaCo-2	0.0
Placenta	0.9	Colon cancer tissue	9.0
Uterus Pool	2.6	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	1.7	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	4.6
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	2.6
Ovarian ca. IGROV-1	8.2	Stomach Pool	3.5
Ovarian ca. OVCAR-8	3.5	Bone Marrow Pool	2.5
Ovary	1.1	Fetal Heart	10.7
Breast ca. MCF-7	0.0	Heart Pool	2.5
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	3.5
Breast ca. BT 549	1.3	Fetal Skeletal Muscle	2.2
Breast ca. T47D	0.0	Skeletal Muscle Pool	1.6
Breast ca. MDA-N	0.1	Spleen Pool	46.7
Breast Pool	4.7	Thymus Pool	2.4
Trachea	2.3	CNS cancer (glio/astro) U87-MG	7.3
Lung	0.5	CNS cancer (glio/astro) U-118-MG	1.3
Fetal Lung	100.0	CNS cancer (neuro;met) SK-N-AS	2.0



Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.3
Lung ca. NCI-H146	1.9	CNS cancer (glio) SNB-19	6.2
Lung ca. SHP-77	8.0	CNS cancer (glio) SF-295	0.1
Lung ca. A549	1.4	Brain (Amygdala) Pool	14.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	9.7
Lung ca. NCI-H23	1.0	Brain (fetal)	34.2
Lung ca. NCI-H460	2.0	Brain (Hippocampus) Pool	15.5
Lung ca. HOP-62	0.4	Cerebral Cortex Pool	22.8
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	21.3
Liver	0.2	Brain (Thalamus) Pool	28.1
Fetal Liver	4.9	Brain (whole)	18.7
Liver ca. HepG2	0.0	Spinal Cord Pool	12.9
Kidney Pool	4.6	Adrenal Gland	2.5
Fetal Kidney	11.3	Pituitary gland Pool	0.7
Renal ca. 786-0	0.0	Salivary Gland	0.3
Renal ca. A498	0.0	Thyroid (female)	1.6
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	9.7

**Table DD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4366, Run 186473465	TISSUE NAME	Rel. Exp.(%) Ag4366, Run 186473465
Secondary Th1 act	0.0	HUVEC IL-1beta	66.0
Secondary Th2 act	0.0	HUVEC IFN gamma	73.7
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	97.9
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	41.5
Secondary Th2 rest	0.0	HUVEC IL-11	13.6
Secondary Tr1 rest	0.0	Lung Microvascular EC none	16.7
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	2.7
Primary Th2 act	0.0	Microvascular Dermal EC none	4.5
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.7	Coronary artery SMC rest	7.6

CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	20.9
CD8 lymphocyte act	0.0	Astrocytes rest	0.7
Secondary CD8 lymphocyte rest	0.5	Astrocytes TNFalpha + IL-1beta	0.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.5
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	1.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	24.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.1
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.5
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	7.6
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	17.1
Two Way MLR 7 day	0.0	Lung fibroblast none	1.1
PBMC rest	0.4	Lung fibroblast TNF alpha + IL-1 beta	0.3
PBMC PWM	0.0	Lung fibroblast IL-4	0.6
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.6
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.4
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.4
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.4
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.5
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.5
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.8
Dendritic cells LPS	0.4	Dermal Fibroblasts rest	1.2
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	1.6
Monocytes LPS	0.0	Colon	4.0
Macrophages rest	0.0	Lung	100.0
Macrophages LPS	0.0	Thymus	9.4
HUVEC none	4.5	Kidney	62.0

HUVEC starved	38.4		
---------------	------	--	--

**Table DE. Panel CNS\_1.1**

Tissue Name	Rel. Exp.(%) Ag4366, Run 195308643	Tissue Name	Rel. Exp.(%) Ag4366, Run 195308643
Cing Gyr Depression2	20.0	BA17 PSP2	3.8
Cing Gyr Depression	9.9	BA17 PSP	13.6
Cing Gyr PSP2	6.6	BA17 Huntington's2	8.3
Cing Gyr PSP	19.3	BA17 Huntington's	14.7
Cing Gyr Huntington's2	30.4	BA17 Parkinson's2	23.2
Cing Gyr Huntington's	73.7	BA17 Parkinson's	26.1
Cing Gyr Parkinson's2	47.6	BA17 Alzheimer's2	5.5
Cing Gyr Parkinson's	49.3	BA17 Control2	28.7
Cing Gyr Alzheimer's2	20.3	BA17 Control	26.2
Cing Gyr Alzheimer's	28.3	BA9 Depression2	6.1
Cing Gyr Control2	54.7	BA9 Depression	6.6
Cing Gyr Control	100.0	BA9 PSP2	4.9
Temp Pole Depression2	10.4	BA9 PSP	11.7
Temp Pole PSP2	4.9	BA9 Huntington's2	12.9
Temp Pole PSP	4.9	BA9 Huntington's	27.2
Temp Pole Huntington's	31.9	BA9 Parkinson's2	53.6
Temp Pole Parkinson's2	26.1	BA9 Parkinson's	16.6
Temp Pole Parkinson's	22.5	BA9 Alzheimer's2	10.8
Temp Pole Alzheimer's2	11.0	BA9 Alzheimer's	5.8
Temp Pole Alzheimer's	5.5	BA9 Control2	54.3
Temp Pole Control2	51.1	BA9 Control	3.2
Temp Pole Control	13.2	BA7 Depression	9.2
Glob Palladus Depression	6.4	BA7 PSP2	11.8
Glob Palladus PSP2	9.3	BA7 PSP	26.4
Glob Palladus PSP	2.6	BA7 Huntington's2	22.5
Glob Palladus Parkinson's2	9.9	BA7 Huntington's	25.3
Glob Palladus Parkinson's	72.2	BA7 Parkinson's2	13.0
Glob Palladus Alzheimer's2	9.6	BA7 Parkinson's	17.9
Glob Palladus Alzheimer's	19.3	BA7 Alzheimer's2	4.8
Glob Palladus Control2	13.1	BA7 Control2	27.4
Glob Palladus Control	20.6	BA7 Control	25.9
Sub Nigra Depression2	13.0	BA4 Depression2	7.1
Sub Nigra Depression	13.1	BA4 Depression	20.0
Sub Nigra PSP2	8.8	BA4 PSP2	30.4

Sub Nigra Huntington's2	31.9	BA4 PSP	6.1
Sub Nigra Huntington's	56.6	BA4 Huntington's2	10.6
Sub Nigra Parkinson's2	67.8	BA4 Huntington's	20.9
Sub Nigra Alzheimer's2	16.0	BA4 Parkinson's2	52.5
Sub Nigra Control2	24.3	BA4 Parkinson's	51.1
Sub Nigra Control	43.2	BA4 Alzheimer's2	8.9
BA17 Depression2	13.1	BA4 Control2	47.0
BA17 Depression	10.6	BA4 Control	30.1

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4366 This panel confirms the expression of the CG108624-01 gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag4366 Highest expression of the CG108624-01 gene is detected in fetal lung (CT=25). Interestingly, expression of this gene is higher in fetal when compared to adult lung (CT=32.7). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal lung suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene may be useful in treatment of muscle related diseases.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. The CG108624-01 gene codes for protocadherin 68. Protocadherins are transmembrane glycoproteins belonging to the cadherin superfamily of molecules, which are involved in many biological processes such as cell adhesion, cytoskeletal organization and morphogenesis. Protocadherins generally exhibit only moderate adhesive activity and are highly expressed in the nervous system. Cadherins can act as axon guidance and cell adhesion proteins, specifically during development and in the response to injury (Ranscht B.,2000 Cadherins: molecular codes for axon guidance and synapse formation. Int. J. Dev. Neurosci. 18: 643-651, PMID: 10978842). Therefore, therapeutic modulation of the levels of this protein may be of use in inducing a compensatory synaptogenic response to neuronal death in Alzheimer's disease, Parkinson's disease, Huntington's disease, spinocerebellar ataxia, progressive supranuclear

palsy, ALS, head trauma, stroke, or any other disease/condition associated with neuronal loss.

Moderate expression of this gene is also seen in samples derived from colon cancer and number of cancer cell lines including brain, lung, breast, ovarian and melanoma cancer  
5 cell lines. Therefore, therapeutic modulation of this gene product may be useful in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the  
10 activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4.1D Summary:** Ag4366 Highest expression of the CG108624-01 gene is detected in lung (CT=28.7). In addition, high expression of this gene is also seen in endothelial cells including HUVEC, HPAEC, lung microvascular and microvascular  
15 dermal endothelial cells. Endothelial cells are known to play important roles in inflammatory responses by altering the expression of surface proteins that are involved in activation and recruitment of effector inflammatory cells. The expression of this gene in these endothelial cells suggests that this protein product may be involved in inflammatory responses to skin and lung disorders, including psoriasis, asthma, allergies, chronic  
20 obstructive pulmonary disease, and emphysema. Therefore, therapeutic modulation of the protein encoded by this gene may lead to amelioration of symptoms associated with psoriasis, asthma, allergies, chronic obstructive pulmonary disease, and emphysema.

Moderate levels of expression of this gene are also seen in resting neutrophils, coronary artery SMC, liver cirrhosis, and normal tissues represented by colon, thymus, and  
25 kidney. Therefore, therapeutic modulation of this gene product may be useful in the treatment of inflammatory and autoimmune diseases that affect colon and kidney including inflammatory bowel diseases, lupus and glomerulonephritis.

**Panel CNS\_1.1 Summary:** Ag4366 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Please see Panel 1.4  
30 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**E. CG108771-01: Type Ib membrane protein**

Expression of full length physical clone CG108771-01 was assessed using the primer-probe set Ag6806, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB, EC and ED.

5

**Table EA. Probe Name Ag6806**

Primers	SEQUENCES	LENGTH	Start Position	SEQ ID No
Forward	5'-cggtgacaggaactgcaaa-3'	19	255	163
Probe	TET-5'-ccctcagggactcggcaaaggctatc-3'-TAMRA	26	275	164
Reverse	5'-agcctggcaggcatga-3'	16	322	165

**Table EB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag6806, Run 278022724	Tissue Name	Rel. Exp.(%) Ag6806, Run 278022724
AD 1 Hippo	19.5	Control (Path) 3 Temporal Ctx	9.9
AD 2 Hippo	53.6	Control (Path) 4 Temporal Ctx	32.3
AD 3 Hippo	18.9	AD 1 Occipital Ctx	30.6
AD 4 Hippo	12.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	90.1	AD 3 Occipital Ctx	13.0
AD 6 Hippo	71.2	AD 4 Occipital Ctx	23.2
Control 2 Hippo	7.0	AD 5 Occipital Ctx	63.3
Control 4 Hippo	20.2	AD 6 Occipital Ctx	25.2
Control (Path) 3 Hippo	9.0	Control 1 Occipital Ctx	10.4
AD 1 Temporal Ctx	28.1	Control 2 Occipital Ctx	75.8
AD 2 Temporal Ctx	44.4	Control 3 Occipital Ctx	28.7
AD 3 Temporal Ctx	15.0	Control 4 Occipital Ctx	12.9
AD 4 Temporal Ctx	25.5	Control (Path) 1 Occipital Ctx	97.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	15.6
AD 5 Sup Temporal Ctx	63.3	Control (Path) 3 Occipital Ctx	7.2
AD 6 Inf Temporal Ctx	66.0	Control (Path) 4 Occipital Ctx	22.8
AD 6 Sup Temporal Ctx	59.9	Control 1 Parietal Ctx	0.6
Control 1 Temporal Ctx	9.7	Control 2 Parietal Ctx	59.0
Control 2 Temporal Ctx	63.7	Control 3 Parietal Ctx	25.7
Control 3 Temporal Ctx	28.7	Control (Path) 1 Parietal Ctx	62.0
Control 3 Temporal Ctx	13.6	Control (Path) 2 Parietal Ctx	28.1
Control (Path) 1 Temporal Ctx	64.6	Control (Path) 3 Parietal Ctx	13.2
Control (Path) 2 Temporal Ctx	57.0	Control (Path) 4 Parietal Ctx	35.4

**Table EC. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6806, Run 278017589	Tissue Name	Rel. Exp.(%) Ag6806, Run 278017589
Adipose	12.3	Renal ca. TK-10	37.9
Melanoma* Hs688(A).T	19.5	Bladder	22.5
Melanoma* Hs688(B).T	18.0	Gastric ca. (liver met.) NCI-N87	30.1
Melanoma* M14	7.8	Gastric ca. KATO III	12.2
Melanoma* LOXIMVI	6.8	Colon ca. SW-948	2.0
Melanoma* SK-MEL-5	5.3	Colon ca. SW480	25.9
Squamous cell carcinoma SCC-4	5.3	Colon ca.* (SW480 met) SW620	9.1
Testis Pool	28.1	Colon ca. HT29	24.3
Prostate ca.* (bone met) PC-3	39.8	Colon ca. HCT-116	21.5
Prostate Pool	18.0	Colon ca. CaCo-2	10.7
Placenta	11.0	Colon cancer tissue	4.9
Uterus Pool	17.9	Colon ca. SW1116	2.7
Ovarian ca. OVCAR-3	10.9	Colon ca. Colo-205	1.8
Ovarian ca. SK-OV-3	39.2	Colon ca. SW-48	0.2
Ovarian ca. OVCAR-4	1.7	Colon Pool	31.0
Ovarian ca. OVCAR-5	56.3	Small Intestine Pool	40.3
Ovarian ca. IGROV-1	9.4	Stomach Pool	21.2
Ovarian ca. OVCAR-8	23.2	Bone Marrow Pool	12.9
Ovary	17.0	Fetal Heart	13.4
Breast ca. MCF-7	15.1	Heart Pool	12.0
Breast ca. MDA-MB-231	21.2	Lymph Node Pool	35.8
Breast ca. BT 549	50.7	Fetal Skeletal Muscle	13.0
Breast ca. T47D	1.7	Skeletal Muscle Pool	1.8
Breast ca. MDA-N	2.4	Spleen Pool	12.1
Breast Pool	37.6	Thymus Pool	27.2
Trachea	20.2	CNS cancer (glio/astro) U87-MG	37.6
Lung	12.7	CNS cancer (glio/astro) U-118-MG	21.2
Fetal Lung	64.6	CNS cancer (neuro;met) SK-N-AS	14.3
Lung ca. NCI-N417	1.2	CNS cancer (astro) SF-539	29.9
Lung ca. LX-1	28.9	CNS cancer (astro) SNB-75	59.0
Lung ca. NCI-H146	5.5	CNS cancer (glio) SNB-19	13.0
Lung ca. SHP-77	18.2	CNS cancer (glio) SF-295	97.9
Lung ca. A549	53.6	Brain (Amygdala) Pool	21.2
Lung ca. NCI-H526	1.7	Brain (cerebellum)	29.1
Lung ca. NCI-H23	53.2	Brain (fetal)	59.9

Lung ca. NCI-H460	7.7	Brain (Hippocampus) Pool	28.1
Lung ca. HOP-62	46.7	Cerebral Cortex Pool	24.1
Lung ca. NCI-H522	100.0	Brain (Substantia nigra) Pool	25.3
Liver	2.0	Brain (Thalamus) Pool	31.2
Fetal Liver	19.9	Brain (whole)	25.3
Liver ca. HepG2	29.9	Spinal Cord Pool	18.6
Kidney Pool	6.3	Adrenal Gland	25.5
Fetal Kidney	93.3	Pituitary gland Pool	12.3
Renal ca. 786-0	35.1	Salivary Gland	6.7
Renal ca. A498	25.2	Thyroid (female)	10.5
Renal ca. ACHN	13.3	Pancreatic ca. CAPAN2	12.2
Renal ca. UO-31	20.7	Pancreas Pool	11.7

**Table ED. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag6806, Run 278020697	TISSUE NAME	Rel. Exp.(%) Ag6806, Run 278020697
Secondary Th1 act	5.3	HUVEC IL-1beta	19.8
Secondary Th2 act	22.7	HUVEC IFN gamma	23.2
Secondary Tr1 act	3.9	HUVEC TNF alpha + IFN gamma	12.9
Secondary Th1 rest	3.2	HUVEC TNF alpha + IL4	6.5
Secondary Th2 rest	3.3	HUVEC IL-11	10.4
Secondary Tr1 rest	2.9	Lung Microvascular EC none	66.4
Primary Th1 act	1.6	Lung Microvascular EC TNFalpha + IL-1beta	12.5
Primary Th2 act	11.8	Microvascular Dermal EC none	6.7
Primary Tr1 act	6.4	Microvascular Dermal EC TNFalpha + IL-1beta	3.6
Primary Th1 rest	0.9	Bronchial epithelium TNFalpha + IL1beta	3.7
Primary Th2 rest	1.9	Small airway epithelium none	4.5
Primary Tr1 rest	1.6	Small airway epithelium TNFalpha + IL-1beta	6.5
CD45RA CD4 lymphocyte act	8.7	Coronary artery SMC rest	11.3
CD45RO CD4 lymphocyte act	9.9	Coronary artery SMC TNFalpha + IL-1beta	15.2
CD8 lymphocyte act	2.8	Astrocytes rest	3.8
Secondary CD8 lymphocyte rest	1.3	Astrocytes TNFalpha + IL-1beta	0.9
Secondary CD8 lymphocyte act	2.3	KU-812 (Basophil) rest	24.0
CD4 lymphocyte none	2.0	KU-812 (Basophil) PMA/ionomycin	18.7
2ry Th1/Th2/Tr1_anti-CD95	4.2	CCD1106 (Keratinocytes) none	6.3



CH11			
LAK cells rest	14.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.6
LAK cells IL-2	2.2	Liver cirrhosis	5.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	2.0
LAK cells IL-2+IFN gamma	1.2	NCI-H292 IL-4	2.1
LAK cells IL-2+ IL-18	1.7	NCI-H292 IL-9	12.1
LAK cells PMA/ionomycin	15.3	NCI-H292 IL-13	6.1
NK Cells IL-2 rest	23.3	NCI-H292 IFN gamma	1.5
Two Way MLR 3 day	6.9	HPAEC none	9.7
Two Way MLR 5 day	2.0	HPAEC TNF alpha + IL-1 beta	22.1
Two Way MLR 7 day	3.6	Lung fibroblast none	9.7
PBMC rest	3.7	Lung fibroblast TNF alpha + IL-1 beta	9.4
PBMC PWM	2.8	Lung fibroblast IL-4	6.4
PBMC PHA-L	3.0	Lung fibroblast IL-9	1.4
Ramos (B cell) none	2.6	Lung fibroblast IL-13	3.0
Ramos (B cell) ionomycin	4.8	Lung fibroblast IFN gamma	18.0
B lymphocytes PWM	0.2	Dermal fibroblast CCD1070 rest	9.5
B lymphocytes CD40L and IL- 4	9.4	Dermal fibroblast CCD1070 TNF alpha	17.1
EOL-1 dbcAMP	21.5	Dermal fibroblast CCD1070 IL-1 beta	4.3
EOL-1 dbcAMP PMA/ionomycin	4.3	Dermal fibroblast IFN gamma	7.1
Dendritic cells none	32.1	Dermal fibroblast IL-4	2.8
Dendritic cells LPS	20.9	Dermal Fibroblasts rest	6.4
Dendritic cells anti-CD40	27.0	Neutrophils TNFa+LPS	13.0
Monocytes rest	18.9	Neutrophils rest	100.0
Monocytes LPS	4.5	Colon	1.5
Macrophages rest	27.2	Lung	0.8
Macrophages LPS	7.9	Thymus	3.2
HUVEC none	11.7	Kidney	14.6
HUVEC starved	25.2		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6806 This panel confirms the expression of this gene at low to moderate levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.6 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.6 Summary:** Ag6806 Expression of the CG108771-01 gene is highest in a lung cancer cell line (CT = 28.2). This gene is expressed at low to moderate levels in the majority of the tissues on this panel. Interestingly, expression of this gene is higher in fetal skeletal muscle, kidney and liver when compared to the adult tissues. Therefore, expression of this gene could be used to distinguish between adult and fetal skeletal muscle, kidney and liver. In addition, the relative overexpression of this gene in fetal skeletal muscle, kidney and liver suggests that the protein product may enhance growth and development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the single-pass transmembrane protein encoded by this gene could be useful in treatment of muscle, kidney and liver related diseases.

This gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4.1D Summary:** Ag6806 Expression of the CG108771-01 gene is highest in resting neutrophils (CT = 29.3), with lower expression detected in activated neutrophils (CT = 32.3). Therefore, expression of this gene could be used to distinguish resting and activated neutrophils. In addition, the CG108771-01 gene product may reduce activation of these inflammatory cells and be useful as a protein therapeutic to reduce or eliminate the symptoms in patients with Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis. In addition, small molecule or antibody antagonists of this gene product may be effective in increasing the immune response in patients with AIDS or other immunodeficiencies.

This gene is also expressed at low to moderate levels in a number of cell types of significance in the immune response in health and disease. These cells include endothelial cells, macrophages/monocytes, basophils, eosinophils, peripheral blood mononuclear cells,

lung and skin epithelial cells, lung and skin fibroblast cells, as well as normal tissues represented by thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in Panel 1.6 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

#### F. CG108782-01 and CG108782-02: Transmembrane Protein

Expression of gene CG108782-01 and full length physical clone CG108782-02 was assessed using the primer-probe sets Ag4367 and Ag6790, described in Tables FA and FB. Results of the RTQ-PCR runs are shown in Tables FC, FD, FE and FF. Please note that CG108782-02 corresponds to Ag6790 only.

**Table FA. Probe Name Ag4367**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - ctgctcactggcttcctctt - 3'	20	803	166
Probe	TET- 5' - ctggcaccaggacgctttgattacat - 3' - TAMRA	26	845	167
Reverse	5' - ggaataactgggtggctgtga - 3'	20	874	168

**Table FB. Probe Name Ag6790**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - gcccgagccgataaaaga - 3'	18	711	169
Probe	TET- 5' - cctatccattcctgttcgacaacctccc - 3' - TAMRA	28	681	170
Reverse	5' - gccttgggctcagtaaggt - 3'	19	642	171

**Table FC. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4367, Run 224376003	Rel. Exp.(%) Ag6790, Run 277731711	Tissue Name	Rel. Exp.(%) Ag4367, Run 224376003	Rel. Exp.(%) Ag6790, Run 277731711
AD 1 Hippo	15.8	20.6	Control (Path) 3 Temporal Ctx	22.7	30.4
AD 2 Hippo	51.1	44.8	Control (Path) 4 Temporal Ctx	33.7	33.9

AD 3 Hippo	10.7	14.4	AD 1 Occipital Ctx	8.5	15.3
AD 4 Hippo	13.0	13.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	41.8	44.1	AD 3 Occipital Ctx	16.4	22.1
AD 6 Hippo	48.6	48.6	AD 4 Occipital Ctx	21.9	27.5
Control 2 Hippo	30.8	60.3	AD 5 Occipital Ctx	43.2	54.3
Control 4 Hippo	34.4	46.0	AD 6 Occipital Ctx	34.9	47.3
Control (Path) 3 Hippo	28.7	33.7	Control 1 Occipital Ctx	24.7	33.2
AD 1 Temporal Ctx	12.1	18.6	Control 2 Occipital Ctx	57.8	75.8
AD 2 Temporal Ctx	30.8	41.8	Control 3 Occipital Ctx	29.5	3.5
AD 3 Temporal Ctx	9.3	13.1	Control 4 Occipital Ctx	34.6	42.6
AD 4 Temporal Ctx	20.4	38.7	Control (Path) 1 Occipital Ctx	100.0	100.0
AD 5 Inf Temporal Ctx	82.9	57.0	Control (Path) 2 Occipital Ctx	28.7	26.8
AD 5 SupTemporal Ctx	49.3	31.6	Control (Path) 3 Occipital Ctx	55.1	70.2
AD 6 Inf Temporal Ctx	44.1	32.1	Control (Path) 4 Occipital Ctx	34.6	26.4
AD 6 Sup Temporal Ctx	35.1	34.6	Control 1 Parietal Ctx	28.3	43.8
Control 1 Temporal Ctx	20.0	36.6	Control 2 Parietal Ctx	34.2	47.6
Control 2 Temporal Ctx	48.6	96.6	Control 3 Parietal Ctx	37.6	49.7
Control 3 Temporal Ctx	20.9	42.3	Control (Path) 1 Parietal Ctx	47.0	80.1
Control 4 Temporal Ctx	28.3	30.1	Control (Path) 2 Parietal Ctx	59.9	71.2
Control (Path) 1 Temporal Ctx	48.6	50.3	Control (Path) 3 Parietal Ctx	46.7	49.0
Control (Path) 2 Temporal Ctx	29.7	46.0	Control (Path) 4 Parietal Ctx	49.7	63.7

**Table FD. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4367, Run 222543793	Tissue Name	Rel. Exp.(%) Ag4367, Run 222543793
Adipose	0.0	Renal ca. TK-10	2.0
Melanoma* Hs688(A).T	0.0	Bladder	0.3
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	2.7
Melanoma* M14	0.7	Gastric ca. KATO III	0.2
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.1	Colon ca. SW480	0.8
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.6
Testis Pool	0.1	Colon ca. HT29	0.6

Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.7
Prostate Pool	0.1	Colon ca. CaCo-2	0.7
Placenta	0.0	Colon cancer tissue	0.3
Uterus Pool	0.1	Colon ca. SW1116	0.5
Ovarian ca. OVCAR-3	0.8	Colon ca. Colo-205	0.7
Ovarian ca. SK-OV-3	0.5	Colon ca. SW-48	0.2
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	5.3	Small Intestine Pool	0.1
Ovarian ca. IGROV-1	2.7	Stomach Pool	0.2
Ovarian ca. OVCAR-8	2.2	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	0.1
Breast ca. MCF-7	1.0	Heart Pool	0.1
Breast ca. MDA-MB-231	0.4	Lymph Node Pool	0.1
Breast ca. BT 549	0.6	Fetal Skeletal Muscle	0.2
Breast ca. T47D	9.3	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.8	Spleen Pool	0.2
Breast Pool	0.0	Thymus Pool	0.2
Trachea	0.2	CNS cancer (glio/astro) U87-MG	0.2
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.4
Fetal Lung	0.3	CNS cancer (neuro;met) SK-N-AS	5.0
Lung ca. NCI-N417	0.1	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	2.6	CNS cancer (astro) SNB-75	3.3
Lung ca. NCI-H146	0.3	CNS cancer (glio) SNB-19	1.8
Lung ca. SHP-77	0.2	CNS cancer (glio) SF-295	4.3
Lung ca. A549	1.2	Brain (Amygdala) Pool	39.2
Lung ca. NCI-H526	0.2	Brain (cerebellum)	12.2
Lung ca. NCI-H23	1.4	Brain (fetal)	1.2
Lung ca. NCI-H460	0.4	Brain (Hippocampus) Pool	26.8
Lung ca. HOP-62	0.3	Cerebral Cortex Pool	24.7
Lung ca. NCI-H522	1.5	Brain (Substantia nigra) Pool	51.8
Liver	0.0	Brain (Thalamus) Pool	30.1
Fetal Liver	0.2	Brain (whole)	8.0
Liver ca. HepG2	0.4	Spinal Cord Pool	100.0
Kidney Pool	0.2	Adrenal Gland	0.0
Fetal Kidney	0.3	Pituitary gland Pool	0.2
Renal ca. 786-0	0.3	Salivary Gland	0.0
Renal ca. A498	0.3	Thyroid (female)	0.0
Renal ca. ACHN	0.3	Pancreatic ca. CAPAN2	1.1
Renal ca. UO-31	0.5	Pancreas Pool	0.2

**Table FE. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6790, Run 277640794	Tissue Name	Rel. Exp.(%) Ag6790, Run 277640794
Adipose	0.0	Renal ca. TK-10	0.6
Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.1	Gastric ca. (liver met.) NCI-N87	2.1
Melanoma* M14	0.3	Gastric ca. KATO III	0.1
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.2	Colon ca. SW480	0.4
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.3
Testis Pool	0.1	Colon ca. HT29	0.4
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.6
Prostate Pool	0.2	Colon ca. CaCo-2	0.3
Placenta	0.1	Colon cancer tissue	0.1
Uterus Pool	0.0	Colon ca. SW1116	0.1
Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	0.6
Ovarian ca. SK-OV-3	0.2	Colon ca. SW-48	0.1
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.0
Ovarian ca. OVCAR-5	2.8	Small Intestine Pool	0.1
Ovarian ca. IGROV-1	0.9	Stomach Pool	0.3
Ovarian ca. OVCAR-8	1.7	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	0.2
Breast ca. MCF-7	0.3	Heart Pool	0.0
Breast ca. MDA-MB-231	0.4	Lymph Node Pool	0.1
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	0.1
Breast ca. T47D	0.8	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.5	Spleen Pool	0.1
Breast Pool	0.1	Thymus Pool	0.1
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.1
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	0.1	CNS cancer (neuro;met) SK-N-AS	2.8
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.1
Lung ca. LX-1	1.6	CNS cancer (astro) SNB-75	2.1
Lung ca. NCI-H146	0.3	CNS cancer (glio) SNB-19	1.6
Lung ca. SHP-77	0.1	CNS cancer (glio) SF-295	3.2
Lung ca. A549	1.1	Brain (Amygdala) Pool	30.1
Lung ca. NCI-H526	0.1	Brain (cerebellum)	16.7
Lung ca. NCI-H23	1.2	Brain (fetal)	0.9

Lung ca. NCI-H460	0.4	Brain (Hippocampus) Pool	20.2
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	26.2
Lung ca. NCI-H522	0.6	Brain (Substantia nigra) Pool	39.5
Liver	0.0	Brain (Thalamus) Pool	39.8
Fetal Liver	0.1	Brain (whole)	9.9
Liver ca. HepG2	0.1	Spinal Cord Pool	100.0
Kidney Pool	0.1	Adrenal Gland	0.0
Fetal Kidney	0.1	Pituitary gland Pool	0.1
Renal ca. 786-0	0.2	Salivary Gland	0.0
Renal ca. A498	0.2	Thyroid (female)	0.0
Renal ca. ACHN	0.3	Pancreatic ca. CAPAN2	0.7
Renal ca. UO-31	0.3	Pancreas Pool	0.1

**Table FF. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4367, Run 186473567	Tissue Name	Rel. Exp.(%) Ag4367, Run 186473567
Secondary Th1 act	0.9	HUVEC IL-1beta	0.0
Secondary Th2 act	4.5	HUVEC IFN gamma	2.3
Secondary Tr1 act	1.0	HUVEC TNF alpha + IFN gamma	1.5
Secondary Th1 rest	2.3	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	6.8	HUVEC IL-11	0.0
Secondary Tr1 rest	13.0	Lung Microvascular EC none	10.3
Primary Th1 act	4.6	Lung Microvascular EC TNFalpha + IL-1beta	2.6
Primary Th2 act	13.5	Microvascular Dermal EC none	0.0
Primary Tr1 act	3.3	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	9.0	Bronchial epithelium TNFalpha + IL1beta	9.7
Primary Th2 rest	1.1	Small airway epithelium none	0.0
Primary Tr1 rest	10.6	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	4.2	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.9
CD8 lymphocyte act	4.4	Astrocytes rest	3.8
Secondary CD8 lymphocyte rest	16.2	Astrocytes TNFalpha + IL-1beta	5.7
Secondary CD8 lymphocyte act	11.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95	8.4	CCD1106 (Keratinocytes) none	1.3

CH11			
LAK cells rest	8.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	11.8
LAK cells IL-2	7.6	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	7.0	NCI-H292 none	100.0
LAK cells IL-2+IFN gamma	8.5	NCI-H292 IL-4	63.3
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	40.3
LAK cells PMA/ionomycin	11.4	NCI-H292 IL-13	40.3
NK Cells IL-2 rest	28.3	NCI-H292 IFN gamma	23.0
Two Way MLR 3 day	3.8	HPAEC none	0.8
Two Way MLR 5 day	6.2	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	15.0	Lung fibroblast none	2.6
PBMC rest	1.2	Lung fibroblast TNF alpha + IL-1 beta	3.9
PBMC PWM	10.9	Lung fibroblast IL-4	0.0
PBMC PHA-L	12.8	Lung fibroblast IL-9	2.2
Ramos (B cell) none	1.9	Lung fibroblast IL-13	0.9
Ramos (B cell) ionomycin	3.4	Lung fibroblast IFN gamma	0.6
B lymphocytes PWM	7.9	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL- 4	1.5	Dermal fibroblast CCD1070 TNF alpha	7.4
EOL-1 dbcAMP	4.5	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	6.7	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	8.7	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	4.7	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	10.7	Neutrophils rest	9.9
Monocytes LPS	17.6	Colon	0.6
Macrophages rest	6.1	Lung	0.0
Macrophages LPS	0.0	Thymus	6.9
HUVEC none	1.6	Kidney	13.3
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4367/Ag6790 Two experiments with two different probe and primer sets are in excellent agreements. These results confirm the expression of this gene at moderate levels in the brain in an independent group of individuals. This gene is downregulated in the temporal cortex of Alzheimer's disease patients when compared with non-demented controls ( $p=0.01$  when analyzed by Ancova, estimate of total cDNA loaded per well used as a covariate). Therefore, up-regulation of this



gene or its protein product, or treatment with specific agonists for this receptor may be of use in reversing the dementia, memory loss, and neuronal death associated with this disease.

**General\_screening\_panel\_v1.4 Summary:** Ag4367 This gene appears to be almost exclusively expressed in the samples originating from the nervous system, with highest expression seen in the spinal cord (26.6). High to moderate levels are also seen in the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Low but significant expression is also seen in many of the cancer cell lines on this panel, including samples derived from pancreatic, colon, gastric, renal, lung, breast, ovarian, and melanoma cancers.

**General\_screening\_panel\_v1.6 Summary:** Ag6790 Expression in this panel is in agreement with expression in Panel 1.4, with highest expression in spinal cord (CT=25.8). High to moderate levels of expression are seen in all CNS regions examined, with low but significant levels of expression in most of the cancer cell lines on this panel. Please see Panel 1.4 for discussion of utility of this gene in central nervous disorders.

**Panel 4.1D Summary:** Ag4367 Highest expression of this gene is seen in an untreated sample from the NCI-H292 pulmonary mucoepidermoid cell line. Lower levels of expression are detected in a cluster of cytokine activated NCI-H292 samples. Thus, the protein could be used to identify certain lung tumors similar to NCI-H292. The encoded protein may also contribute to the normal function of the goblet cells within the lung. Therefore, designing therapeutics to this protein may be important for the treatment of emphysema and asthma as well as other lung diseases in which goblet cells or the mucus they produce have pathological consequences. A second experiment with Ag6790 showed low/undetectable levels of expression. (CTs>35). (Data not shown.)

#### **G. CG108801-01 and CG108801-02: EGF-domain Transmembrane Protein**

Expression of gene CG108801-01 and variant CG108801-02 was assessed using the primer-probe sets Ag2449 and Ag737, described in Tables GA and GB. Results of the RTQ-PCR runs are shown in Tables GC, GD, GE and GF.

**Table GA. Probe Name Ag2449**

Primers	Sequences	Length	Start	SEQ ID
---------	-----------	--------	-------	--------

			Position	No
Forward	5' - accacagcactgtgatgga - 3'	19	1322	172
Probe	TET-5' - acctcttccaccctcagactggagt - 3' - TAMRA	26	1368	173
Reverse	5' - agctgaagggtggtgagaac - 3'	20	1399	174

**Table GB. Probe Name Ag737**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - gcaagacctgtgagcttgctc - 3'	20	915	175
Probe	TET-5' - caaccacagtggacacccctctagg - 3' - TAMRA	25	954	176
Reverse	5' - cgtggcaggtaccactacag - 3'	20	990	177

**Table GC. Panel 1.3D**

Tissue Name	Rel. Exp.(%) Ag2449, Run 160661517	Rel. Exp.(%) Ag2449, Run 165630571	Tissue Name	Rel. Exp.(%) Ag2449, Run 160661517	Rel. Exp.(%) Ag2449, Run 165630571
Liver adenocarcinoma	9.4	13.5	Kidney (fetal)	1.4	5.6
Pancreas	2.1	3.4	Renal ca. 786-0	0.4	13.0
Pancreatic ca. CAPAN 2	1.3	4.9	Renal ca. A498	3.4	46.7
Adrenal gland	1.3	0.6	Renal ca. RXF 393	12.3	40.3
Thyroid	2.2	4.3	Renal ca. ACHN	0.0	7.1
Salivary gland	7.3	9.7	Renal ca. UO-31	3.9	16.3
Pituitary gland	3.5	2.1	Renal ca. TK-10	0.0	7.3
Brain (fetal)	16.0	4.2	Liver	0.5	0.6
Brain (whole)	26.6	2.7	Liver (fetal)	0.8	8.8
Brain (amygdala)	19.5	3.4	Liver ca. (hepatoblast) HepG2	43.2	2.0
Brain (cerebellum)	10.5	0.4	Lung	2.4	38.2
Brain (hippocampus)	100.0	6.1	Lung (fetal)	3.5	53.6
Brain (substantia nigra)	3.4	0.7	Lung ca. (small cell) LX-1	2.9	25.2
Brain (thalamus)	9.4	4.4	Lung ca. (small cell) NCI-H69	53.2	2.3
Cerebral Cortex	27.7	2.6	Lung ca. (s.cell var.) SHP-77	7.6	0.0
Spinal cord	0.5	1.1	Lung ca. (large cell)NCI-H460	1.3	2.3
glio/astro U87-MG	0.5	0.0	Lung ca. (non-sm. cell) A549	8.0	1.6
glio/astro U-118-MG	9.0	100.0	Lung ca. (non-s.cell) NCI-H23	31.2	4.2
astrocytoma SW1783	4.5	44.1	Lung ca. (non-s.cell)	4.4	8.8

			HOP-62		
neuro*; met SK-N-AS	3.9	0.0	Lung ca. (non-s.cl) NCI-H522	51.4	2.9
astrocytoma SF-539	5.0	4.3	Lung ca. (squam.) SW 900	0.9	10.5
astrocytoma SNB-75	2.2	31.2	Lung ca. (squam.) NCI-H596	2.6	1.1
glioma SNB-19	2.8	1.7	Mammary gland	24.8	8.2
glioma U251	3.7	6.9	Breast ca.* (pl.ef) MCF-7	2.0	4.2
glioma SF-295	1.2	15.4	Breast ca.* (pl.ef) MDA-MB-231	63.7	8.7
Heart (fetal)	17.3	8.2	Breast ca.* (pl.ef) T47D	0.5	6.3
Heart	1.9	8.8	Breast ca. BT-549	27.7	94.6
Skeletal muscle (fetal)	10.8	5.7	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.0	7.3	Ovary	0.6	3.8
Bone marrow	1.3	7.2	Ovarian ca. OVCAR-3	4.1	10.7
Thymus	6.0	7.3	Ovarian ca. OVCAR-4	0.3	1.8
Spleen	2.1	9.2	Ovarian ca. OVCAR-5	3.1	8.0
Lymph node	4.4	34.2	Ovarian ca. OVCAR-8	0.6	6.0
Colorectal	1.8	0.0	Ovarian ca. IGROV-1	0.0	12.7
Stomach	5.6	16.6	Ovarian ca.* (ascites) SK-OV-3	8.1	9.7
Small intestine	5.2	13.0	Uterus	1.4	16.3
Colon ca. SW480	12.6	9.0	Placenta	1.8	2.7
Colon ca.* SW620(SW480 met)	2.1	1.9	Prostate	28.7	15.7
Colon ca. HT29	2.6	0.6	Prostate ca.* (bone met)PC-3	5.8	4.5
Colon ca. HCT-116	2.0	3.3	Testis	2.6	1.9
Colon ca. CaCo-2	2.0	1.1	Melanoma Hs688(A).T	0.2	55.5
Colon ca. tissue(ODO3866)	1.3	12.3	Melanoma* (met) Hs688(B).T	0.0	37.6
Colon ca. HCC-2998	11.1	10.9	Melanoma UACC-62	0.5	1.6
Gastric ca.* (liver met) NCI-N87	2.7	70.7	Melanoma M14	0.4	5.0
Bladder	0.5	12.2	Melanoma LOX IMVI	4.8	1.0
Trachea	49.7	21.6	Melanoma* (met) SK- MEL-5	1.2	0.0
Kidney	1.1	1.8	Adipose	0.4	14.1

**Table GD. Panel 2D**

Tissue Name	Rel. Exp.(%) Ag2449, Run 160661561	Tissue Name	Rel. Exp.(%) Ag2449, Run 160661561
Normal Colon	19.3	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	0.8	Kidney Cancer 8120613	3.4
CC Margin (ODO3866)	5.6	Kidney Margin 8120614	7.0
CC Gr.2 rectosigmoid (ODO3868)	4.9	Kidney Cancer 9010320	7.2
CC Margin (ODO3868)	2.9	Kidney Margin 9010321	9.2
CC Mod Diff (ODO3920)	11.9	Normal Uterus	0.5
CC Margin (ODO3920)	2.8	Uterus Cancer 064011	6.4
CC Gr.2 ascend colon (ODO3921)	3.1	Normal Thyroid	0.3
CC Margin (ODO3921)	1.2	Thyroid Cancer 064010	1.4
CC from Partial Hepatectomy (ODO4309) Mets	1.5	Thyroid Cancer A302152	2.5
Liver Margin (ODO4309)	0.4	Thyroid Margin A302153	1.9
Colon mets to lung (OD04451-01)	0.0	Normal Breast	8.9
Lung Margin (OD04451-02)	0.7	Breast Cancer (OD04566)	4.8
Normal Prostate 6546-1	40.1	Breast Cancer (OD04590-01)	5.3
Prostate Cancer (OD04410)	15.0	Breast Cancer Mets (OD04590-03)	3.9
Prostate Margin (OD04410)	43.8	Breast Cancer Metastasis (OD04655-05)	1.8
Prostate Cancer (OD04720-01)	75.3	Breast Cancer 064006	5.8
Prostate Margin (OD04720-02)	100.0	Breast Cancer 1024	37.1
Normal Lung 061010	12.3	Breast Cancer 9100266	8.8
Lung Met to Muscle (ODO4286)	3.2	Breast Margin 9100265	10.6
Muscle Margin (ODO4286)	0.6	Breast Cancer A209073	14.9
Lung Malignant Cancer (OD03126)	3.7	Breast Margin A209073	16.3
Lung Margin (OD03126)	4.9	Normal Liver	0.0
Lung Cancer (OD04404)	18.0	Liver Cancer 064003	2.5
Lung Margin (OD04404)	12.9	Liver Cancer 1025	1.2
Lung Cancer (OD04565)	12.9	Liver Cancer 1026	2.6
Lung Margin (OD04565)	2.8	Liver Cancer 6004-T	6.1
Lung Cancer (OD04237-01)	4.3	Liver Tissue 6004-N	2.3
Lung Margin (OD04237-02)	2.6	Liver Cancer 6005-T	5.1
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	1.5
Melanoma Mets to Lung (OD04321)	5.1	Bladder Cancer 1023	0.0

Lung Margin (OD04321)	4.3	Bladder Cancer A302173	4.5
Normal Kidney	6.3	Bladder Cancer (OD04718-01)	39.0
Kidney Ca, Nuclear grade 2 (OD04338)	1.4	Bladder Normal Adjacent (OD04718-03)	1.0
Kidney Margin (OD04338)	1.7	Normal Ovary	2.3
Kidney Ca Nuclear grade 1/2 (OD04339)	3.2	Ovarian Cancer 064008	10.4
Kidney Margin (OD04339)	1.2	Ovarian Cancer (OD04768-07)	11.0
Kidney Ca, Clear cell type (OD04340)	4.2	Ovary Margin (OD04768-08)	4.3
Kidney Margin (OD04340)	3.1	Normal Stomach	1.9
Kidney Ca, Nuclear grade 3 (OD04348)	11.2	Gastric Cancer 9060358	0.1
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	1.0
Kidney Cancer (OD04622-01)	1.4	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	1.2	Stomach Margin 9060394	1.6
Kidney Cancer (OD04450-01)	3.2	Gastric Cancer 9060397	2.7
Kidney Margin (OD04450-03)	0.6	Stomach Margin 9060396	0.8
Kidney Cancer 8120607	0.7	Gastric Cancer 064005	3.1

**Table GE. Panel 3D**

Tissue Name	Rel. Exp.(%) Ag2449, Run 164827286	Tissue Name	Rel. Exp.(%) Ag2449, Run 164827286
Daoy- Medulloblastoma	1.1	Ca Ski- Cervical epidermoid carcinoma (metastasis)	81.2
TE671- Medulloblastoma	0.8	ES-2- Ovarian clear cell carcinoma	0.3
D283 Med- Medulloblastoma	11.0	Ramos- Stimulated with PMA/ionomycin 6h	0.6
PFSK-1- Primitive Neuroectodermal	2.7	Ramos- Stimulated with PMA/ionomycin 14h	0.6
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megakaryoblast)	1.3
SNB-78- Glioma	0.6	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	4.3	Daudi- Burkitt's lymphoma	0.7
T98G- Glioblastoma	5.4	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	2.5	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	1.5	RL- non-Hodgkin's B-cell lymphoma	0.3
Cerebellum	8.6	JM1- pre-B-cell lymphoma	0.6
Cerebellum	12.7	Jurkat- T cell leukemia	2.3

NCI-H292- Mucoepidermoid lung carcinoma	20.9	TF-1- Erythroleukemia	2.1
DMS-114- Small cell lung cancer	18.0	HUT 78- T-cell lymphoma	3.0
DMS-79- Small cell lung cancer	67.8	U937- Histiocytic lymphoma	2.1
NCI-H146- Small cell lung cancer	16.8	KU-812- Myelogenous leukemia	1.7
NCI-H526- Small cell lung cancer	3.4	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	5.4	Caki-2- Clear cell renal carcinoma	0.4
NCI-H82- Small cell lung cancer	4.4	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	41.5	G401- Wilms' tumor	1.8
NCI-H1155- Large cell lung cancer	28.9	Hs766T- Pancreatic carcinoma (LN metastasis)	2.0
NCI-H1299- Large cell lung cancer	13.5	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.7
NCI-H727- Lung carcinoid	0.4	SU86.86- Pancreatic carcinoma (liver metastasis)	6.4
NCI-UMC-11- Lung carcinoid	5.3	BxPC-3- Pancreatic adenocarcinoma	38.7
LX-1- Small cell lung cancer	1.7	HPAC- Pancreatic adenocarcinoma	2.2
Colo-205- Colon cancer	1.1	MIA PaCa-2- Pancreatic carcinoma	5.7
KM12- Colon cancer	4.9	CFPAC-1- Pancreatic ductal adenocarcinoma	3.8
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	32.5
NCI-H716- Colon cancer	10.7	T24- Bladder carcinoma (transitional cell)	1.7
SW-48- Colon adenocarcinoma	0.3	5637- Bladder carcinoma	2.1
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	1.2
LS 174T- Colon adenocarcinoma	1.0	UM-UC-3- Bladder carcinoma (transitional cell)	2.3
SW-948- Colon adenocarcinoma	0.3	A204- Rhabdomyosarcoma	7.3
SW-480- Colon adenocarcinoma	0.3	HT-1080- Fibrosarcoma	3.8
NCI-SNU-5- Gastric carcinoma	9.6	MG-63- Osteosarcoma	0.2
KATO III- Gastric carcinoma	2.5	SK-LMS-1- Leiomyosarcoma (vulva)	3.4
NCI-SNU-16- Gastric carcinoma	0.6	SJRH30- Rhabdomyosarcoma (met to bone marrow)	2.1
NCI-SNU-1- Gastric carcinoma	2.5	A431- Epidermoid carcinoma	100.0

RF-1- Gastric adenocarcinoma	4.7	WM266-4- Melanoma	1.2
RF-48- Gastric adenocarcinoma	5.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	1.3	MDA-MB-468- Breast adenocarcinoma	2.3
NCI-N87- Gastric carcinoma	0.7	SCC-4- Squamous cell carcinoma of tongue	0.3
OVCAR-5- Ovarian carcinoma	0.2	SCC-9- Squamous cell carcinoma of tongue	0.4
RL95-2- Uterine carcinoma	13.6	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	1.4	CAL 27- Squamous cell carcinoma of tongue	17.6

**Table GF. Panel 4D**

Tissue Name	Rel. Exp.(%) Ag2449, Run 159516095	Rel. Exp.(%) Ag2449, Run 162622351	Tissue Name	Rel. Exp.(%) Ag2449, Run 159516095	Rel. Exp.(%) Ag2449, Run 162622351
Secondary Th1 act	6.4	2.3	HUVEC IL-1beta	1.5	3.1
Secondary Th2 act	5.2	8.1	HUVEC IFN gamma	3.8	2.5
Secondary Tr1 act	5.4	6.2	HUVEC TNF alpha + IFN gamma	0.9	0.6
Secondary Th1 rest	0.7	1.8	HUVEC TNF alpha + IL4	0.7	2.4
Secondary Th2 rest	0.8	2.4	HUVEC IL-11	2.5	2.8
Secondary Tr1 rest	2.8	3.4	Lung Microvascular EC none	3.2	2.1
Primary Th1 act	5.0	4.2	Lung Microvascular EC TNFalpha + IL- 1beta	1.0	1.1
Primary Th2 act	8.9	1.8	Microvascular Dermal EC none	5.4	1.1
Primary Tr1 act	7.0	8.6	Microvascular Dermal EC TNFalpha + IL- 1beta	1.2	2.5
Primary Th1 rest	4.8	5.9	Bronchial epithelium TNFalpha + IL1beta	0.0	1.0
Primary Th2 rest	3.8	3.2	Small airway epithelium none	6.0	2.9
Primary Tr1 rest	3.1	5.3	Small airway epithelium TNFalpha + IL- 1beta	35.6	33.4

CD45RA CD4 lymphocyte act	4.7	0.6	Coronary artery SMC rest	1.4	1.3
CD45RO CD4 lymphocyte act	7.4	5.3	Coronary artery SMC TNFalpha + IL-1beta	1.0	0.0
CD8 lymphocyte act	2.6	4.2	Astrocytes rest	0.6	0.7
Secondary CD8 lymphocyte rest	12.2	8.1	Astrocytes TNFalpha + IL-1beta	1.7	2.0
Secondary CD8 lymphocyte act	4.8	2.1	KU-812 (Basophil) rest	9.5	2.9
CD4 lymphocyte none	1.3	2.9	KU-812 (Basophil) PMA/ionomycin	9.5	4.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.6	2.6	CCD1106 (Keratinocytes) none	90.1	80.1
LAK cells rest	2.0	4.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.1	6.2
LAK cells IL-2	10.7	5.4	Liver cirrhosis	4.5	0.0
LAK cells IL-2+IL-12	6.7	5.7	Lupus kidney	0.0	1.3
LAK cells IL-2+IFN gamma	8.8	4.7	NCI-H292 none	100.0	100.0
LAK cells IL-2+ IL-18	10.2	4.5	NCI-H292 IL-4	71.7	95.3
LAK cells PMA/ionomycin	4.6	1.6	NCI-H292 IL-9	97.3	100.0
NK Cells IL-2 rest	6.6	5.4	NCI-H292 IL-13	57.4	47.6
Two Way MLR 3 day	6.7	6.6	NCI-H292 IFN gamma	46.7	39.5
Two Way MLR 5 day	10.4	3.3	HPAEC none	1.9	2.1
Two Way MLR 7 day	1.7	1.1	HPAEC TNF alpha + IL-1 beta	4.1	1.5
PBMC rest	0.0	0.0	Lung fibroblast none	3.8	2.7
PBMC PWM	12.8	15.4	Lung fibroblast TNF alpha + IL-1 beta	0.1	0.0
PBMC PHA-L	10.7	9.0	Lung fibroblast IL-4	8.0	6.5
Ramos (B cell) none	10.9	3.2	Lung fibroblast IL-9	5.1	4.5
Ramos (B cell) ionomycin	10.2	17.3	Lung fibroblast IL-13	2.3	1.3
B lymphocytes	8.3	9.9	Lung fibroblast IFN	1.8	4.2



PWM			gamma		
B lymphocytes CD40L and IL-4	10.4	4.1	Dermal fibroblast CCD1070 rest	5.8	4.7
EOL-1 dbcAMP	4.0	6.0	Dermal fibroblast CCD1070 TNF alpha	13.2	8.0
EOL-1 dbcAMP PMA/ionomycin	13.2	7.3	Dermal fibroblast CCD1070 IL-1 beta	0.6	2.9
Dendritic cells none	1.3	4.5	Dermal fibroblast IFN gamma	2.2	2.2
Dendritic cells LPS	3.4	4.4	Dermal fibroblast IL-4	7.1	5.1
Dendritic cells anti- CD40	2.7	2.2	IBD Colitis 2	0.6	0.0
Monocytes rest	4.2	2.5	IBD Crohn's	0.8	0.7
Monocytes LPS	5.1	3.1	Colon	24.3	7.6
Macrophages rest	2.7	1.9	Lung	5.1	5.7
Macrophages LPS	1.3	0.0	Thymus	2.9	6.9
HUVEC none	1.1	1.9	Kidney	25.9	23.0
HUVEC starved	1.3	0.7			

**Panel 1.3D Summary:** Ag2449 Results from two experiments using the same probe-primer set were in poor agreement and no conclusions can be made (data not shown).

**Panel 2D Summary:** Ag2449 Expression of the CG108801-01 gene is highest in the prostate-derived samples, both normal and cancerous (CTs = 30-32). Therefore, expression of this gene could be used to distinguish prostate from the other tissues on this panel. In addition, this gene may play a role in normal prostate function. This gene is also expressed at low levels in normal colon, breast, kidney and lung.

**Panel 3D Summary:** Ag2449 Expression of the CG108801-01 gene is highest in an epidermoid carcinoma cell line (CT = 28). Moderate expression of this gene is also detected in a subset of pancreatic cancer and lung cancer cell lines. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of these types of cancer.

**Panel 4D Summary:** Ag2449 Results from two experiments using the same probe-primer set gave results that are in good agreement. Expression of the CG108801-01 gene is highest in the NCI-H292 cell line, a human airway epithelial cell line that produces mucins (CT = 29). Mucus overproduction is an important feature of bronchial asthma and chronic obstructive pulmonary disease. The transcript is also expressed at lower but significant

levels in small airway epithelium treated with IL-1 beta and TNF-alpha. The expression of the transcript in this mucoepidermoid cell line that is often used as a model for airway epithelium (NCI-H292 cells) suggests that this transcript may be important in the proliferation or activation of airway epithelium. Therefore, therapeutics designed with the protein encoded by the transcript may reduce or eliminate symptoms caused by inflammation in lung epithelia in chronic obstructive pulmonary disease, asthma, allergy, and emphysema.

This gene is also expressed at moderate levels in resting keratinocytes (CT = 29.3) and at lower levels in treated keratinocytes (CT = 32.6). Therefore, modulation of the expression or activity of the protein encoded by this transcript through the application of small molecule therapeutics may be useful in the treatment of psoriasis and wound healing.

#### H. CG109717-01: Transmembrane Protein

Expression of gene CG109717-01 was assessed using the primer-probe sets Ag4296 and Ag4396, described in Tables HA and HB. Results of the RTQ-PCR runs are shown in Tables HC, HD and HE.

**Table HA. Probe Name Ag4296**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -acctgaaagaactggaggaaaa-3'	22	2384	178
Probe	TET-5' -tcactagcattcattctgtggccttg-3' -TAMRA	26	2431	179
Reverse	5' -gttcttggctcactgaagtcac-3'	22	2457	180

**Table HB. Probe Name Ag4396**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -acctgaaagaactggaggaaaa-3'	22	2384	181
Probe	TET-5' -tcactagcattcattctgtggccttg-3' -TAMRA	26	2431	182
Reverse	5' -gttcttggctcactgaagtcac-3'	22	2457	183

**Table HC. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4296, Run 224073755	Rel. Exp.(%) Ag4396, Run 224504220	Tissue Name	Rel. Exp.(%) Ag4296, Run 224073755	Rel. Exp.(%) Ag4396, Run 224504220
AD 1 Hippo	32.3	27.0	Control (Path) 3 Temporal Ctx	3.7	5.6
AD 2 Hippo	48.3	58.2	Control (Path) 4 Temporal Ctx	62.9	89.5
AD 3 Hippo	26.8	16.0	AD 1 Occipital Ctx	25.0	1.2
AD 4 Hippo	20.6	19.1	AD 2 Occipital Ctx	0.0	0.0

			(Missing)		
AD 5 hippo	89.5	70.7	AD 3 Occipital Ctx	6.5	5.0
AD 6 Hippo	46.7	51.1	AD 4 Occipital Ctx	26.4	25.7
Control 2 Hippo	45.4	33.4	AD 5 Occipital Ctx	11.4	34.6
Control 4 Hippo	7.0	10.4	AD 6 Occipital Ctx	52.1	9.4
Control (Path) 3 Hippo	5.6	7.5	Control 1 Occipital Ctx	1.2	0.6
AD 1 Temporal Ctx	25.5	21.9	Control 2 Occipital Ctx	36.1	30.6
AD 2 Temporal Ctx	44.4	43.2	Control 3 Occipital Ctx	19.2	27.7
AD 3 Temporal Ctx	17.6	15.1	Control 4 Occipital Ctx	4.0	3.1
AD 4 Temporal Ctx	38.4	29.9	Control (Path) 1 Occipital Ctx	87.7	76.3
AD 5 Inf Temporal Ctx	88.9	74.7	Control (Path) 2 Occipital Ctx	24.3	21.8
AD 5 SupTemporal Ctx	54.3	77.9	Control (Path) 3 Occipital Ctx	0.4	1.8
AD 6 Inf Temporal Ctx	45.4	46.7	Control (Path) 4 Occipital Ctx	25.7	26.1
AD 6 Sup Temporal Ctx	42.3	40.9	Control 1 Parietal Ctx	5.8	5.0
Control 1 Temporal Ctx	5.2	5.5	Control 2 Parietal Ctx	55.5	41.8
Control 2 Temporal Ctx	36.3	25.5	Control 3 Parietal Ctx	23.2	17.1
Control 3 Temporal Ctx	38.2	32.8	Control (Path) 1 Parietal Ctx	82.4	68.3
Control 4 Temporal Ctx	14.0	12.2	Control (Path) 2 Parietal Ctx	31.4	40.1
Control (Path) 1 Temporal Ctx	100.0	100.0	Control (Path) 3 Parietal Ctx	2.6	2.8
Control (Path) 2 Temporal Ctx	85.9	67.4	Control (Path) 4 Parietal Ctx	80.7	80.1

**Table HD. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4296, Run 222184254	Rel. Exp.(%) Ag4396, Run 222641544	Tissue Name	Rel. Exp.(%) Ag4296, Run 222184254	Rel. Exp.(%) Ag4396, Run 222641544
Adipose	0.7	0.3	Renal ca. TK-10	0.6	1.2
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.1	0.0
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0

Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	1.7	0.9	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	1.3	1.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.0	0.0
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.2	0.7	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	0.0	0.2
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	0.2	0.1
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.5	0.0
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	0.0	0.0
Ovary	2.0	2.3	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.0	0.0
Breast ca. MDA-MB-231	0.0	0.0	Lymph Node Pool	0.3	0.6
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	0.0	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	0.0	0.0	Spleen Pool	1.5	1.3
Breast Pool	0.0	0.1	Thymus Pool	0.1	0.2
Trachea	0.1	0.2	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.1	0.0	CNS cancer (glio/astro) U-118-MG	0.3	0.0
Fetal Lung	0.0	0.0	CNS cancer (neuro;met) SK-N-AS	0.3	0.2
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.2	0.1
Lung ca. NCI-H146	2.0	2.2	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	2.4	0.6	CNS cancer (glio) SF-295	1.1	0.3
Lung ca. A549	0.0	0.0	Brain (Amygdala) Pool	37.1	31.6
Lung ca. NCI-H526	0.3	0.0	Brain (cerebellum)	6.5	10.9
Lung ca. NCI-H23	0.1	0.2	Brain (fetal)	100.0	100.0
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus)	59.9	50.0

			Pool		
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	90.1	78.5
Lung ca. NCI-H522	0.0	0.3	Brain (Substantia nigra) Pool	42.9	36.3
Liver	0.0	0.0	Brain (Thalamus) Pool	92.0	82.9
Fetal Liver	0.0	0.3	Brain (whole)	74.7	65.1
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	11.0	6.3
Kidney Pool	0.0	0.0	Adrenal Gland	1.5	1.3
Fetal Kidney	0.0	0.2	Pituitary gland Pool	1.7	1.0
Renal ca. 786-0	0.0	0.0	Salivary Gland	0.1	0.0
Renal ca. A498	0.0	0.2	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.5	0.0

**Table HE. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4296, Run 181981947	Rel. Exp.(%) Ag4396, Run 187513671	Tissue Name	Rel. Exp.(%) Ag4296, Run 181981947	Rel. Exp.(%) Ag4396, Run 187513671
Secondary Th1 act	16.7	15.6	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	14.0	18.2	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	24.3	13.7	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	3.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	5.4	Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0
Primary Th1 rest	3.5	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	0.0
Primary Tr1 rest	6.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	0.0
CD45RA CD4	10.2	3.7	Coronary artery SMC	0.0	0.0

lymphocyte act			rest		
CD45RO CD4 lymphocyte act	18.9	24.5	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	67.4	40.3	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	18.2	20.4	Astrocytes TNFalpha + IL-1beta	0.0	0.0
Secondary CD8 lymphocyte act	26.2	8.1	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	4.7	KU-812 (Basophil) PMA/ionomycin	0.0	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	3.7	15.9	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	22.8	18.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0
LAK cells IL-2	100.0	100.0	Liver cirrhosis	0.0	4.7
LAK cells IL-2+IL-12	28.5	18.6	NCI-H292 none	0.0	0.0
LAK cells IL-2+IFN gamma	23.2	11.5	NCI-H292 IL-4	0.0	0.0
LAK cells IL-2+ IL-18	28.9	40.6	NCI-H292 IL-9	0.0	0.0
LAK cells PMA/ionomycin	0.0	8.6	NCI-H292 IL-13	0.0	0.0
NK Cells IL-2 rest	29.9	74.2	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 3 day	14.5	23.2	HPAEC none	0.0	0.0
Two Way MLR 5 day	27.4	26.2	HPAEC TNF alpha + IL-1 beta	0.0	0.0
Two Way MLR 7 day	11.3	15.6	Lung fibroblast none	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0
PBMC PWM	8.4	9.3	Lung fibroblast IL-4	0.0	0.0
PBMC PHA-L	16.7	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes PWM	4.8	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
B lymphocytes CD40L and IL-4	11.0	4.2	Dermal fibroblast CCD1070 TNF alpha	29.3	33.0
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IL-4	0.0	0.0
Dendritic cells LPS	0.0	0.0	Dermal Fibroblasts rest	0.0	0.0

Dendritic cells anti-CD40	0.0	0.0	Neutrophils TNFa+LPS	0.0	0.0
Monocytes rest	0.0	0.0	Neutrophils rest	4.6	3.4
Monocytes LPS	3.6	0.0	Colon	0.0	0.0
Macrophages rest	0.0	0.0	Lung	0.0	4.0
Macrophages LPS	0.0	0.0	Thymus	0.0	0.0
HUVEC none	0.0	0.0	Kidney	0.0	0.0
HUVEC starved	0.0	0.0			

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag4396 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4396/Ag4296 Two experiments with same probe and primer sets are in excellent agreement. This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly down-regulated in the temporal cortex of Alzheimer's disease patients. Therefore, up-regulation of this gene or its protein product, or treatment with specific agonists for this receptor may be of use in reversing the dementia, memory loss, and neuronal death associated with this disease.

**General\_screening\_panel\_v1.4 Summary:** Ag4396/Ag4296 Two experiments with same probe and primer sets are in excellent agreement, with highest expression of this gene in fetal brain (CTs=29). High expression of this gene is seen in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, expression of this gene may be used to differentiate brain samples from other samples used in this panel. Furthermore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

In addition, low levels of expression of this gene are also seen in two lung cancer cell lines. Therefore, therapeutic modulation of this gene may be useful in the treatment of lung cancer.

**Panel 4.1D Summary:** Ag4396/Ag4296 Two experiments with same probe and primer sets are in excellent agreement, with highest expression of this gene in IL-2 treated LAK cells. Therefore, expression of this gene may be used to distinguish this sample from other samples used in this panel. Low levels of expression of this gene is also seen in IL-2 treated NK cells. These killer cells are involved in tumor immunology and cell clearance of virally and bacterial infected cells as well as tumors. Therefore, modulation of the function of the protein encoded by this gene through the application of a small molecule drug or

antibody may alter the functions of these cells and lead to improvement of symptoms associated with these conditions.

### I. CG110477-01: Desmoglein 3 variant

Expression of gene CG110477-01 was assessed using the primer-probe set Ag4420,  
5 described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB, IC and ID.

**Table IA. Probe Name Ag4420**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - tttcaatgacaattgtccaaca - 3'	22	1464	184
Probe	TET-5' - cagtttgagtttcttcaccttcctg - 3' - TAMRA	26	1505	185
Reverse	5' - attcagtgttcttagcggagaca - 3'	22	1533	186

**Table IB. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4420, Run 219925591	Tissue Name	Rel. Exp.(%) Ag4420, Run 219925591
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	100.0
Melanoma* M14	0.0	Gastric ca. KATO III	14.8
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	28.5	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.2	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.3	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	14.7
Uterus Pool	26.6	Colon ca. SW1116	0.7
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	58.6
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	52.9
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	0.0



Breast ca. T47D	0.1	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.1	Thymus Pool	9.2
Trachea	11.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	2.1	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.5	Brain (fetal)	0.3
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.3
Liver	0.0	Brain (Thalamus) Pool	0.1
Fetal Liver	0.0	Brain (whole)	0.7
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.1	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.8
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**Table IC. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4420, Run 190281898	Tissue Name	Rel. Exp.(%) Ag4420, Run 190281898
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.1
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0

Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	100.0
Primary Th2 rest	0.0	Small airway epithelium none	47.6
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	95.3
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.4
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	1.7
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	27.5
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	51.8
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	35.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	53.2
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	59.9
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	47.3
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	49.3
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.1	Lung fibroblast TNF alpha + IL-1 beta	0.1
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.1
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.1	Dermal fibroblast CCD1070 IL-1 beta	0.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.6
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.2
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.1
Monocytes rest	0.0	Neutrophils rest	0.6
Monocytes LPS	0.0	Colon	1.5

Macrophages rest	0.0	Lung	2.0
Macrophages LPS	0.0	Thymus	25.3
HUVEC none	0.0	Kidney	23.3
HUVEC starved	0.0		

**Table ID. general oncology screening panel\_v\_2.4**

Tissue Name	Rel. Exp.(%) Ag4420, Run 268665926	Tissue Name	Rel. Exp.(%) Ag4420, Run 268665926
Colon cancer 1	1.1	Bladder NAT 2	0.0
Colon NAT 1	0.0	Bladder NAT 3	1.0
Colon cancer 2	19.9	Bladder NAT 4	0.0
Colon NAT 2	0.0	Prostate adenocarcinoma 1	1.0
Colon cancer 3	1.0	Prostate adenocarcinoma 2	0.0
Colon NAT 3	0.1	Prostate adenocarcinoma 3	0.3
Colon malignant cancer 4	51.8	Prostate adenocarcinoma 4	7.9
Colon NAT 4	0.0	Prostate NAT 5	0.3
Lung cancer 1	0.1	Prostate adenocarcinoma 6	0.0
Lung NAT 1	0.0	Prostate adenocarcinoma 7	0.0
Lung cancer 2	0.0	Prostate adenocarcinoma 8	0.0
Lung NAT 2	0.0	Prostate adenocarcinoma 9	0.3
Squamous cell carcinoma 3	100.0	Prostate NAT 10	0.0
Lung NAT 3	0.0	Kidney cancer 1	0.0
Metastatic melanoma 1	0.0	Kidney NAT 1	0.0
Melanoma 2	6.0	Kidney cancer 2	0.0
Melanoma 3	4.3	Kidney NAT 2	0.0
Metastatic melanoma 4	0.0	Kidney cancer 3	0.0
Metastatic melanoma 5	0.1	Kidney NAT 3	0.0
Bladder cancer 1	0.0	Kidney cancer 4	0.0
Bladder NAT 1	0.0	Kidney NAT 4	0.0
Bladder cancer 2	0.7		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4420 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.4 Summary:** Ag4420 Expression of this gene is restricted to a few samples on this panel, with highest expression in a gastric cancer cell line (CT=25.4). High levels of expression are also seen in a colon cancer and squamous cell carcinoma cell lines. Moderate levels of expression are seen in thymus, uterus, and trachea, with low but significant levels detected in cerebral cortex, substantia nigra, and fetal and

whole brain samples. Thus, expression of this gene could be used to differentiate the gastric cancer cell line from other samples on this panel and as a marker of gastric cancer. This gene product is homologous to desmoglein, a component of intercellular desmosome junctions, involved in mediating cell-cell adhesion. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of gastric, colon or skin cancer.

**Panel 4.1D Summary:** Ag4420 Expression on this panel appears to be limited to a few samples, with highest expression in TNF- $\alpha$  and IL1- $\beta$  activated bronchial and small airway epithelium (CTs=27.2). Moderate levels of expression are also seen in untreated small airway epithelium, treated and untreated keratinocytes, a cluster of samples derived from NCI-H292 cells, and thymus and kidney. This expression profile suggests that this gene product may be involved in inflammatory conditions of the lung and skin. Modulation of the expression or function of this gene may be useful in the treatment of psoriasis, asthma, allergy and emphysema.

**general oncology screening panel\_v\_2.4 Summary:** Ag4420 Highest expression of this gene in this panel is seen in a squamous cell carcinoma sample (CT=25.4). High levels of expression are also seen in samples from colon cancer, with moderate levels of expression in melanoma and prostate cancer. This expression is in agreement with the expression seen in cancer cell lines in panel 1.4. Thus, expression of this gene could be used as a marker of these cancers. Modulation of the expression or function of this gene may be useful in the treatment of skin, colon or prostate cancer.

#### J. CG110540-01: pheromone receptor

Expression of gene CG110540-01 was assessed using the primer-probe set Ag4422, described in Table JA. Results of the RTQ-PCR runs are shown in Tables JB, JC, JD and JE.

**Table JA. Probe Name Ag4422**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aagaaatgctgctttctctgaa-3'	22	43	187
Probe	TET-5'-atctctgccaatgccatgctcctt-3'-TAMRA	24	77	188
Reverse	5'-cacgtgaggatgtggaagag-3'	20	101	189

**Table JB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4422, Run 224502250	Tissue Name	Rel. Exp.(%) Ag4422, Run 224502250
-------------	---------------------------------------------	-------------	------------------------------------------

	Run 224502250		
AD 1 Hippo	9.3	Control (Path) 3 Temporal Ctx	4.0
AD 2 Hippo	43.5	Control (Path) 4 Temporal Ctx	29.9
AD 3 Hippo	7.5	AD 1 Occipital Ctx	17.1
AD 4 Hippo	4.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	81.8	AD 3 Occipital Ctx	3.3
AD 6 Hippo	20.6	AD 4 Occipital Ctx	3.8
Control 2 Hippo	25.3	AD 5 Occipital Ctx	12.6
Control 4 Hippo	7.0	AD 6 Occipital Ctx	20.2
Control (Path) 3 Hippo	5.9	Control 1 Occipital Ctx	1.5
AD 1 Temporal Ctx	9.5	Control 2 Occipital Ctx	21.3
AD 2 Temporal Ctx	44.4	Control 3 Occipital Ctx	9.1
AD 3 Temporal Ctx	5.1	Control 4 Occipital Ctx	4.8
AD 4 Temporal Ctx	14.8	Control (Path) 1 Occipital Ctx	87.1
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	10.0
AD 5 Sup Temporal Ctx	52.9	Control (Path) 3 Occipital Ctx	0.8
AD 6 Inf Temporal Ctx	26.8	Control (Path) 4 Occipital Ctx	8.4
AD 6 Sup Temporal Ctx	23.8	Control 1 Parietal Ctx	6.0
Control 1 Temporal Ctx	5.1	Control 2 Parietal Ctx	59.5
Control 2 Temporal Ctx	29.3	Control 3 Parietal Ctx	7.3
Control 3 Temporal Ctx	13.7	Control (Path) 1 Parietal Ctx	80.7
Control 4 Temporal Ctx	6.4	Control (Path) 2 Parietal Ctx	9.6
Control (Path) 1 Temporal Ctx	65.5	Control (Path) 3 Parietal Ctx	2.4
Control (Path) 2 Temporal Ctx	27.4	Control (Path) 4 Parietal Ctx	29.9

**Table JC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4422, Run 219925646	Tissue Name	Rel. Exp.(%) Ag4422, Run 219925646
Adipose	3.3	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	6.7	Bladder	5.3
Melanoma* Hs688(B).T	2.5	Gastric ca. (liver met.) NCI-N87	6.8
Melanoma* M14	0.0	Gastric ca. KATO III	11.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	3.3
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	9.9
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	1.6
Testis Pool	74.2	Colon ca. HT29	1.2
Prostate ca.* (bone met) PC-3	26.1	Colon ca. HCT-116	4.6

Prostate Pool	11.0	Colon ca. CaCo-2	2.2
Placenta	7.4	Colon cancer tissue	4.0
Uterus Pool	17.7	Colon ca. SW1116	3.5
Ovarian ca. OVCAR-3	4.9	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	7.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.8	Colon Pool	21.6
Ovarian ca. OVCAR-5	1.9	Small Intestine Pool	88.9
Ovarian ca. IGROV-1	5.5	Stomach Pool	10.5
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	13.0
Ovary	6.3	Fetal Heart	1.2
Breast ca. MCF-7	0.4	Heart Pool	12.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	27.9
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	12.5
Breast ca. T47D	2.2	Skeletal Muscle Pool	2.2
Breast ca. MDA-N	1.2	Spleen Pool	12.6
Breast Pool	25.0	Thymus Pool	14.0
Trachea	3.4	CNS cancer (glio/astro) U87-MG	0.0
Lung	7.9	CNS cancer (glio/astro) U-118-MG	4.3
Fetal Lung	14.6	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	1.4
Lung ca. LX-1	35.8	CNS cancer (astro) SNB-75	14.3
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.5
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	3.8
Lung ca. A549	5.0	Brain (Amygdala) Pool	40.1
Lung ca. NCI-H526	1.0	Brain (cerebellum)	34.6
Lung ca. NCI-H23	0.0	Brain (fetal)	100.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	36.6
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	54.7
Lung ca. NCI-H522	24.1	Brain (Substantia nigra) Pool	50.3
Liver	1.1	Brain (Thalamus) Pool	51.4
Fetal Liver	0.0	Brain (whole)	84.1
Liver ca. HepG2	0.0	Spinal Cord Pool	20.2
Kidney Pool	67.4	Adrenal Gland	3.0
Fetal Kidney	23.2	Pituitary gland Pool	1.7
Renal ca. 786-0	0.0	Salivary Gland	0.5
Renal ca. A498	4.9	Thyroid (female)	3.0
Renal ca. ACHN	2.4	Pancreatic ca. CAPAN2	1.3
Renal ca. UO-31	3.4	Pancreas Pool	27.4

**Table JD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4422, Run 190282090	Tissue Name	Rel. Exp.(%) Ag4422, Run 190282090
Secondary Th1 act	0.9	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.8
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.9	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	2.0	Lung Microvascular EC none	1.4
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.5
Primary Th2 act	0.0	Microvascular Dermal EC none	0.1
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.6	Bronchial epithelium TNFalpha + IL1beta	1.4
Primary Th2 rest	0.9	Small airway epithelium none	0.5
Primary Tr1 rest	1.7	Small airway epithelium TNFalpha + IL-1beta	2.3
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.4
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.9	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.9	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.4	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.9	Liver cirrhosis	0.9
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.8	NCI-H292 IL-4	0.4
LAK cells IL-2+ IL-18	3.9	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	2.1	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.5	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.5	Lung fibroblast none	0.0
PBMC rest	4.4	Lung fibroblast TNF alpha + IL-1 beta	1.0

PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.5	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.8
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.4	Dermal fibroblast CCD1070 IL-1 beta	0.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.5
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.5
Dendritic cells anti-CD40	0.5	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.8
Monocytes LPS	0.0	Colon	2.9
Macrophages rest	0.0	Lung	7.1
Macrophages LPS	0.0	Thymus	14.9
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

**Table JE. general oncology screening panel\_v\_2.4**

Tissue Name	Rel. Exp.(%) Ag4422, Run 268665945	Tissue Name	Rel. Exp.(%) Ag4422, Run 268665945
Colon cancer 1	2.0	Bladder NAT 2	0.0
Colon NAT 1	5.1	Bladder NAT 3	2.9
Colon cancer 2	0.7	Bladder NAT 4	15.4
Colon NAT 2	5.7	Prostate adenocarcinoma 1	100.0
Colon cancer 3	10.4	Prostate adenocarcinoma 2	2.0
Colon NAT 3	12.6	Prostate adenocarcinoma 3	36.3
Colon malignant cancer 4	4.9	Prostate adenocarcinoma 4	39.5
Colon NAT 4	0.0	Prostate NAT 5	2.0
Lung cancer 1	5.7	Prostate adenocarcinoma 6	9.0
Lung NAT 1	0.0	Prostate adenocarcinoma 7	8.8
Lung cancer 2	38.2	Prostate adenocarcinoma 8	7.5
Lung NAT 2	1.9	Prostate adenocarcinoma 9	52.1
Squamous cell carcinoma 3	6.3	Prostate NAT 10	6.0
Lung NAT 3	0.0	Kidney cancer 1	9.2
Metastatic melanoma 1	38.7	Kidney NAT 1	13.1
Melanoma 2	0.0	Kidney cancer 2	18.9



Melanoma 3	1.8	Kidney NAT 2	9.8
Metastatic melanoma 4	47.0	Kidney cancer 3	33.9
Metastatic melanoma 5	38.7	Kidney NAT 3	8.2
Bladder cancer 1	1.3	Kidney cancer 4	0.0
Bladder NAT 1	0.0	Kidney NAT 4	1.8
Bladder cancer 2	23.2		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4452 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

5        **General\_screening\_panel\_v1.4 Summary:** Ag4452 Highest expression of this gene is found in the fetal brain (CT=30.2). Prominent expression of this gene is also seen throughout the CNS. This gene encodes a putative member of the pheromone receptor family. These receptors are expressed in sensory neurons and are involved in the initiation of innate reproductive and social behaviors. From the tissue distribution and predicted  
10 function of this gene, expression of this gene could be used to differentiate between brain and non-neuronal tissue. In addition, modulation of the expression or function of this gene may be useful in the treatment of behavioural and reproductive disorders.

Low but significant expression is seen in pancreas, heart, and fetal skeletal muscle, all tissues with metabolic function. This expression suggests that this gene product may  
15 also be involved in neuroendocrine function.

**Panel 4.1D Summary:** Ag4452 Highest expression of this gene is seen in the kidney (CT=29.3), with moderate expression detected in colon, lung, thymus, resting PBMCs, and IL-18/IL-2 treated LAK cells. Thus, expression of this gene could be used to differentiate the kidney derived sample from other samples on this panel and as a marker of  
20 kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

**general oncology screening panel\_v\_2.4 Summary:** Ag4452 Highest expression of this gene is seen in prostate cancer (CT=32.1). Low but significant levels of expression  
25 are also seen in melanoma, lung and kidney cancer.

**K. CG110725-01: OSTEOPONTIN PRECURSOR**

Expression of full length physical clone CG110725-01 was assessed using the primer-probe sets Ag6782 and Ag6796, described in Tables KA and KB. Results of the RTQ-PCR runs are shown in Tables KC and KD.

5

**Table KA. Probe Name Ag6782**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ccgtgaattccacagccat-3'	19	800	190
Probe	TET-5'-caaccagcatatcttcatggctgtgaaattc-3'-TAMRA	31	819	191
Reverse	5'-atcttcttctccttacttttgggggtct-3'	25	851	192

**Table KB. Probe Name Ag6796**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ccacaagcagtcctcaaagtca-3'	20	779	193
Probe	TET-5'-ccgtgaattccacagccatgaatttc-3'-TAMRA	26	800	194
Reverse	5'-ggctctacaaccagcatatcttcat-3'	24	832	195

**Table KC. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag6782, Run 277731702	Rel. Exp.(%) Ag6782, Run 283829326	Tissue Name	Rel. Exp.(%) Ag6782, Run 277731702	Rel. Exp.(%) Ag6782, Run 283829326
AD 1 Hippo	19.2	12.6	Control (Path) 3 Temporal Ctx	0.7	1.1
AD 2 Hippo	19.6	18.0	Control (Path) 4 Temporal Ctx	4.5	3.3
AD 3 Hippo	14.5	11.2	AD 1 Occipital Ctx	22.4	16.2
AD 4 Hippo	5.0	3.9	AD 2 Occipital Ctx (Missing)	0.0	0.1
AD 5 Hippo	50.3	76.3	AD 3 Occipital Ctx	10.8	10.0
AD 6 Hippo	48.3	46.0	AD 4 Occipital Ctx	11.9	11.7
Control 2 Hippo	26.1	19.8	AD 5 Occipital Ctx	15.8	21.5
Control 4 Hippo	13.6	8.7	AD 6 Occipital Ctx	30.8	23.3
Control (Path) 3 Hippo	5.0	4.3	Control 1 Occipital Ctx	1.3	1.0
AD 1 Temporal Ctx	20.3	9.7	Control 2 Occipital Ctx	18.3	20.6
AD 2 Temporal Ctx	30.1	12.6	Control 3 Occipital Ctx	5.8	5.7
AD 3 Temporal Ctx	5.6	4.5	Control 4 Occipital Ctx	7.6	5.4
AD 4 Temporal Ctx	16.4	10.4	Control (Path) 1 Occipital Ctx	27.7	20.9
AD 5 Inf Temporal Ctx	100.0	100.0	Control (Path) 2	3.3	5.2

			Occipital Ctx		
AD 5 Sup Temporal Ctx	42.9	73.7	Control (Path) 3 Occipital Ctx	3.8	3.0
AD 6 Inf Temporal Ctx	41.5	41.8	Control (Path) 4 Occipital Ctx	5.5	6.1
AD 6 Sup Temporal Ctx	35.6	29.9	Control 1 Parietal Ctx	2.7	3.4
Control 1 Temporal Ctx	1.3	0.8	Control 2 Parietal Ctx	40.9	43.2
Control 2 Temporal Ctx	18.7	4.9	Control 3 Parietal Ctx	12.4	7.4
Control 3 Temporal Ctx	6.0	4.2	Control (Path) 1 Parietal Ctx	9.5	11.0
Control 3 Temporal Ctx	3.4	3.0	Control (Path) 2 Parietal Ctx	11.9	11.4
Control (Path) 1 Temporal Ctx	9.5	10.5	Control (Path) 3 Parietal Ctx	1.2	1.3
Control (Path) 2 Temporal Ctx	19.3	15.0	Control (Path) 4 Parietal Ctx	10.9	10.2

**Table KD. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6782, Run 278015361	Rel. Exp.(%) Ag6796, Run 278017527	Tissue Name	Rel. Exp.(%) Ag6782, Run 278015361	Rel. Exp.(%) Ag6796, Run 278017527
Adipose	0.6	0.0	Renal ca. TK-10	4.7	11.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	9.0	11.5
Melanoma* Hs688(B).T	0.1	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	2.7	5.1	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	1.7	2.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	4.5	5.1	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.3	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.2	0.0	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	0.0	0.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.8	5.1	Colon cancer tissue	13.2	27.0
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.1	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	5.3	7.9	Colon ca. SW-48	0.1	0.0
Ovarian ca. OVCAR-4	0.1	0.0	Colon Pool	0.0	0.0
Ovarian ca. OVCAR-5	0.2	0.0	Small Intestine Pool	0.1	0.0
Ovarian ca. IGROV-1	100.0	80.1	Stomach Pool	0.1	0.0
Ovarian ca. OVCAR-8	1.6	2.6	Bone Marrow Pool	0.0	0.0

Ovary	0.1	0.0	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.0	0.0
Breast ca. MDA-MB-231	0.0	0.0	Lymph Node Pool	0.0	0.0
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	0.1	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	4.5	2.4	Spleen Pool	0.0	0.0
Breast Pool	0.0	1.7	Thymus Pool	0.1	0.0
Trachea	0.1	0.0	CNS cancer (glio/astro) U87-MG	8.2	19.6
Lung	0.0	0.0	CNS cancer (glio/astro) U-118-MG	0.0	0.0
Fetal Lung	0.4	0.0	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.1	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	73.2	100.0
Lung ca. SHP-77	2.5	2.7	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. A549	1.8	4.7	Brain (Amygdala) Pool	3.3	3.0
Lung ca. NCI-H526	0.0	0.0	Brain (cerebellum)	1.0	2.5
Lung ca. NCI-H23	0.1	1.1	Brain (fetal)	0.8	11.5
Lung ca. NCI-H460	9.5	43.2	Brain (Hippocampus) Pool	3.4	3.3
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	2.1	15.5
Lung ca. NCI-H522	0.1	0.0	Brain (Substantia nigra) Pool	1.6	11.7
Liver	0.0	0.0	Brain (Thalamus) Pool	3.6	6.6
Fetal Liver	0.1	0.0	Brain (whole)	0.6	3.0
Liver ca. HepG2	0.8	0.0	Spinal Cord Pool	8.2	29.5
Kidney Pool	0.1	0.0	Adrenal Gland	0.0	0.0
Fetal Kidney	0.8	3.0	Pituitary gland Pool	0.0	0.0
Renal ca. 786-0	8.7	15.3	Salivary Gland	0.0	0.0
Renal ca. A498	8.2	18.9	Thyroid (female)	0.0	0.0
Renal ca. ACHN	12.5	26.6	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	2.9	2.0	Pancreas Pool	0.5	4.5

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6782 This panel confirms the expression of the CG110725-01 gene in the brains of an independent group of individuals. This gene appears to be upregulated in the temporal cortex of Alzheimer's disease patients.

Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Ag6796 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

5        **General\_screening\_panel\_v1.6 Summary:** Ag6782 Expression of the CG110725-01 gene is highest in an ovarian cancer cell line (CT = 21.2). This gene is also expressed at higher levels in a subset of lung and renal cancer cell lines when compared to their respective normal tissues. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or  
10        antibodies, might be beneficial in the treatment of lung, renal and ovarian cancer.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy,  
15        multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed in pancreas, adipose, adrenal gland, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity  
20        and diabetes.

The CG110725-01 gene encodes a splice variant form of osteopontin, a protein produced by osteoblasts under stimulation by calcitriol that is involved in the anchoring osteoclasts to the mineral of bone matrix. Osteopontin is one of the key cytokines for type 1 immune responses mediated by macrophages in mice (S. Ashkar et al., Science 287: 860-  
25        864, 2000, PubMed ID : 10657301). In addition, osteopontin has been shown to be overexpressed in a variety of human tumors and is present in elevated levels in the blood of some patients with metastatic cancers (K.A. Furger et al., Curr Mol Med. 2001 Nov;1(5):621-32, PMID: 11899236). The osteopontin protein is thought to play a role in tumor invasion and metastasis through integrin-mediated signal transduction. These  
30        observations suggest that the osteopontin splice variant described here may be useful as a dominant negative in the treatment of cancer.

Ag6796 The pattern of gene expression in this experiment is similar to what is seen with Ag6782, but the levels of expression are much lower. The Ag6796 and Ag6782 probe-primer sets recognize distinct regions of this gene.

**Panel 4.1D Summary:** Ag6782 Results from one experiment with the CG110725-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Ag6796 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

#### **L. CG111683-03: PULMONARY SURFACTANT-ASSOCIATED PROTEIN C PRECURSOR**

Expression of full length physical clone CG111683-03 was assessed using the primer-probe set Ag6780, described in Table LA. Results of the RTQ-PCR runs are shown in Tables LB, LC and LD.

**Table LA. Probe Name Ag6780**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -attgtggaagcccagcaa-3'	18	144	196
Probe	TET-5' -ctgagtgagcacctggttaccactgcc-3' - TAMRA	27	171	197
Reverse	5' -agtggagccgatggagaa-3'	18	201	198

**Table LB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag6780, Run 277731699	Tissue Name	Rel. Exp.(%) Ag6780, Run 277731699
AD 1 Hippo	13.3	Control (Path) 3 Temporal Ctx	8.7
AD 2 Hippo	29.7	Control (Path) 4 Temporal Ctx	25.9
AD 3 Hippo	12.3	AD 1 Occipital Ctx	28.7
AD 4 Hippo	8.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	63.7	AD 3 Occipital Ctx	18.2
AD 6 Hippo	50.0	AD 4 Occipital Ctx	30.1
Control 2 Hippo	56.6	AD 5 Occipital Ctx	45.4
Control 4 Hippo	10.7	AD 6 Occipital Ctx	43.2
Control (Path) 3 Hippo	13.0	Control 1 Occipital Ctx	10.4
AD 1 Temporal Ctx	38.2	Control 2 Occipital Ctx	87.7
AD 2 Temporal Ctx	31.6	Control 3 Occipital Ctx	25.3
AD 3 Temporal Ctx	15.8	Control 4 Occipital Ctx	12.2
AD 4 Temporal Ctx	16.8	Control (Path) 1 Occipital Ctx	70.7
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	13.2
AD 5 Sup Temporal Ctx	20.6	Control (Path) 3 Occipital Ctx	9.5

AD 6 Inf Temporal Ctx	42.6	Control (Path) 4 Occipital Ctx	18.3
AD 6 Sup Temporal Ctx	34.6	Control 1 Parietal Ctx	14.7
Control 1 Temporal Ctx	5.7	Control 2 Parietal Ctx	42.3
Control 2 Temporal Ctx	55.1	Control 3 Parietal Ctx	38.7
Control 3 Temporal Ctx	15.3	Control (Path) 1 Parietal Ctx	64.2
Control 3 Temporal Ctx	14.7	Control (Path) 2 Parietal Ctx	27.0
Control (Path) 1 Temporal Ctx	39.5	Control (Path) 3 Parietal Ctx	9.6
Control (Path) 2 Temporal Ctx	28.5	Control (Path) 4 Parietal Ctx	32.5

**Table LC. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6780, Run 278015357	Tissue Name	Rel. Exp.(%) Ag6780, Run 278015357
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.4	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.2	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.1	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.0

Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	100.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	0.0	Brain (fetal)	0.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.3
Liver	0.0	Brain (Thalamus) Pool	0.3
Fetal Liver	0.0	Brain (whole)	0.2
Liver ca. HepG2	0.0	Spinal Cord Pool	0.5
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**Table LD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag6780, Run 277641299	Tissue Name	Rel. Exp.(%) Ag6780, Run 277641299
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0



Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.1
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	100.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6780 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

**General\_screening\_panel\_v1.6 Summary:** Ag6780 Highest expression of this gene is seen in the fetal lung (CT=22.7). Thus, expression of this gene could be used to differentiate between fetal and adult lung tissue (CT=40). This gene product has homology to the surfactant-associated protein SP-C, a hydrophobic, lung-specific protein, that enhances the surface-tension-lowering properties of surfactant lipids and helps in stabilizing the respiratory surface of lungs against collapse. Noguee et. al demonstrated that alterations in this gene SP-C may be related to the development of pulmonary disease in the adult (N Engl J Med 2001, 344:573579.). Devendra proposed that abnormalities in the lung surfactant system may also play a role in the development of chronic obstructive pulmonary disease, and asthma (Respir Res 2002;3(1):19). Thus, based on the homology of this gene product to SP-C and the highly specific expression seen in the developin lung, modulation of the expression or function of this gene may be useful in the treatment these lung related diseases.

This gene is also expressed at low but significant levels in the brain. Please see CNS\_neurodegeneration\_v1.0 for discussion of utility of this gene in the CNS.

**Panel 4.1D Summary:** Ag6780 Expression of this gene is exclusive to the lung in this panel (CT=26.2). Thus, expression of this gene could be used as a marker of lung tissue. Please see Panel 1.6 for further discussion of utility of this gene in autoimmune disease.

## **M. CG112655-01: GERM CELL-LESS 1 PROTEIN**

Expression of gene CG112655-01 was assessed using the primer-probe set Ag6812, described in Table MA. Results of the RTQ-PCR runs are shown in Table MB.

**Table MA. Probe Name Ag6812**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcgagctacggcttctgtta-3'	20	175	199
Probe	TET-5'-tccgcttgcgcttgtgactgcc-3'-TAMRA	22	202	200
Reverse	5'-gcgagtcggggtcaca-3'	16	244	201

**Table MB. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6812, Run 278018586	Tissue Name	Rel. Exp.(%) Ag6812, Run 278018586
Adipose	1.1	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.9
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.6
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	100.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	5.0
Placenta	0.0	Colon cancer tissue	1.3
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	2.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	6.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	1.5
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	4.6
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.7	Thymus Pool	8.2
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	1.1	CNS cancer (glio/astro) U-118-MG	7.1
Fetal Lung	0.7	CNS cancer (neuro;met) SK-N-AS	0.6
Lung ca. NCI-N417	0.5	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	2.3	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	1.3	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0

Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	2.3
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	5.4
Lung ca. NCI-H522	4.8	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	2.4
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	9.1	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6812 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.6 Summary:** Ag6812 Expression of this gene is exclusive to the testis (CT=31.8). Thus, expression of this gene could be used to

- 5 differentiate between this sample and other samples on this panel and as a marker of testicular tissue. Therapeutic modulation of the expression or function of this gene may be useful in the treatment of male infertility and hypogonadism.

**Panel 4.1D Summary:** Ag6812 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

10 **N. CG112813-01 and CG112813-02 and CG112813-04 and CG112813-05 and CG112813-06: NATURAL KILLER CELL RECEPTOR**

- Expression of gene CG112813-01, variants CG112813-05 and CG112813-06, and full length physical clones CG112813-02 and CG112813-04 was assessed using the primer-probe sets Ag4465, Ag4783, Ag4784, Ag5089, Ag6237, Ag6508, Ag6654 and Ag6247, described in Tables NB, NC, ND, NE, NF, NG, NH, and NI. The correspondence between the individual variants and the probes and primer sets is indicated in Table NA. Results of the RTQ-PCR runs are shown in Tables NJ, NK, NL and NM.
- 15

**Table NA. Correspondence between probe and primer sets and individual sequences**

	Ag4465	Ag4783	Ag4784	Ag5089	Ag6237	Ag6247	Ag6508	Ag6654
CG112813-01	X	X	X	X				
CG112813-02	X	X		X				
CG112813-04	X	X		X				
CG112813-05	X	X	X		X		X	X

CG112813-06		X				X		
-------------	--	---	--	--	--	---	--	--

**Table NB. Probe Name Ag4465**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggacattgtgatcacaggaaaa-3'	22	646	202
Probe	TET-5'-aaaagccttctctctccaccagggtg-3'-TAMRA	26	672	203
Reverse	5'-gcagaagagggtcaacttctct-3'	22	718	204

**Table NC. Probe Name Ag4783**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cagagacacaggaagcattc-3'	21	815	205
Probe	TET-5'-ccagtccctggcgaggacattatag-3'-TAMRA	23	866	206
Reverse	5'-agtcattgaaggaacattatagca-3'	22	890	207

**Table ND. Probe Name Ag4784**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cctctcggaggatcctagtatc-3'	22	1108	208
Probe	TET-5'-tacatcactgtccaccaagccagg-3'-TAMRA	25	1130	209
Reverse	5'-cttgcgatttaattgcctttg-3'	21	1186	210

**Table NE. Probe Name Ag5089**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aactcatcacgcctctgtgtct-3'	22	18	211
Probe	TET-5'-acactggcctttccaacaac-3'-TAMRA	20	222	212
Reverse	5'-ttccaacacatctgttaggtccc-3'	22	275	213

5

**Table NF. Probe Name Ag6508**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acagaatccaccctgaatc-3'	20	978	214
Probe	TET-5'-tgacaccaccatggcaaacacagag-3'-TAMRA	25	998	215
Reverse	5'-ctgcaggctcctcttctc-3'	19	1044	216

**Table NG. Probe Name Ag6654**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acagaatccaccctgaatc-3'	20	978	217
Probe	TET-5'-tgacaccaccatggcaaacacagag-3'-TAMRA	25	998	218
Reverse	5'-ctgcaggctcctcttctc-3'	19	1044	219

**Table NH. Probe Name Ag6237**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcagagacacaggaagcat-3'	20	807	220
Probe	TET-5'-ctctccctataagccccaggt-3'-TAMRA	21	903	221
Reverse	5'-gtgggtgtcagattcaggggt-3'	20	987	222

**Table NJ. AI\_comprehensive panel\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4783, Run 212319919	Rel. Exp.(%) Ag4783, Run 212746557	Tissue Name	Rel. Exp.(%) Ag4783, Run 212319919	Rel. Exp.(%) Ag4783, Run 212746557
110967 COPD-F	0.0	0.0	112427 Match Control Psoriasis-F	0.0	7.1
110980 COPD-F	0.0	0.0	112418 Psoriasis-M	0.0	0.0
110968 COPD-M	0.0	4.8	112723 Match Control Psoriasis-M	0.0	2.3
110977 COPD-M	4.8	0.0	112419 Psoriasis-M	0.0	0.0
110989 Emphysema-F	0.0	0.0	112424 Match Control Psoriasis-M	0.0	1.9
110992 Emphysema-F	12.9	11.0	112420 Psoriasis-M	35.1	24.7
110993 Emphysema-F	4.0	0.0	112425 Match Control Psoriasis-M	0.0	1.9
110994 Emphysema-F	0.0	0.0	104689 (MF) OA Bone- Backus	2.5	5.4
110995 Emphysema-F	48.3	60.3	104690 (MF) Adj "Normal" Bone-Backus	2.2	2.6
110996 Emphysema-F	25.2	20.2	104691 (MF) OA Synovium-Backus	0.0	0.0
110997 Asthma-M	0.0	1.9	104692 (BA) OA Cartilage-Backus	0.0	0.0
111001 Asthma-F	0.0	0.0	104694 (BA) OA Bone- Backus	0.0	0.0
111002 Asthma-F	2.4	0.0	104695 (BA) Adj "Normal" Bone-Backus	0.0	2.6
111003 Atopic Asthma-F	1.0	6.1	104696 (BA) OA Synovium-Backus	1.6	0.0
111004 Atopic Asthma-F	6.7	8.8	104700 (SS) OA Bone- Backus	2.1	2.0
111005 Atopic Asthma-F	2.5	4.3	104701 (SS) Adj "Normal" Bone-Backus	0.0	2.5
111006 Atopic Asthma-F	1.0	0.0	104702 (SS) OA Synovium-Backus	1.5	0.0
111417 Allergy-M	3.1	1.6	117093 OA Cartilage Rep7	13.0	36.1
112347 Allergy-M	0.0	1.5	112672 OA Bone5	5.4	13.6
112349 Normal Lung-F	0.0	1.8	112673 OA Synovium5	4.8	0.0
112357 Normal Lung-F	0.0	0.0	112674 OA Synovial Fluid cells5	1.7	6.6

112354 Normal Lung-M	0.0	6.4	117100 OA Cartilage Rep14	5.3	3.2
112374 Crohns-F	3.2	1.4	112756 OA Bone9	0.0	0.0
112389 Match Control Crohns-F	0.0	0.0	112757 OA Synovium9	0.0	0.0
112375 Crohns-F	0.0	0.0	112758 OA Synovial Fluid Cells9	0.0	0.0
112732 Match Control Crohns-F	18.9	10.7	117125 RA Cartilage Rep2	0.0	0.0
112725 Crohns-M	1.9	0.0	113492 Bone2 RA	8.2	10.9
112387 Match Control Crohns-M	9.0	0.0	113493 Synovium2 RA	0.0	4.8
112378 Crohns-M	2.4	12.4	113494 Syn Fluid Cells RA	4.3	9.3
112390 Match Control Crohns-M	2.4	9.7	113499 Cartilage4 RA	8.3	6.1
112726 Crohns-M	1.5	2.4	113500 Bone4 RA	3.5	12.2
112731 Match Control Crohns-M	0.0	0.0	113501 Synovium4 RA	4.0	9.3
112380 Ulcer Col-F	0.0	2.4	113502 Syn Fluid Cells4 RA	0.0	3.3
112734 Match Control Ulcer Col-F	17.6	16.7	113495 Cartilage3 RA	6.2	3.0
112384 Ulcer Col-F	100.0	100.0	113496 Bone3 RA	8.1	6.3
112737 Match Control Ulcer Col-F	0.0	0.0	113497 Synovium3 RA	7.4	5.6
112386 Ulcer Col-F	0.0	2.2	113498 Syn Fluid Cells3 RA	7.9	3.6
112738 Match Control Ulcer Col-F	1.6	0.0	117106 Normal Cartilage Rep20	0.0	0.0
112381 Ulcer Col-M	0.0	0.0	113663 Bone3 Normal	0.0	0.0
112735 Match Control Ulcer Col-M	0.0	0.0	113664 Synovium3 Normal	0.0	0.0
112382 Ulcer Col-M	0.0	0.0	113665 Syn Fluid Cells3 Normal	0.0	4.0
112394 Match Control Ulcer Col-M	2.3	1.8	117107 Normal Cartilage Rep22	0.0	0.0
112383 Ulcer Col-M	72.2	80.7	113667 Bone4 Normal	11.7	11.9
112736 Match Control Ulcer Col-M	0.0	3.2	113668 Synovium4 Normal	11.3	12.2
112423 Psoriasis-F	1.7	0.0	113669 Syn Fluid Cells4 Normal	19.3	21.2

Table NK. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel.	Tissue Name	Rel.
-------------	------	-------------	------

	Exp.(%) Ag4465, Run 224535063		Exp.(%) Ag4465, Run 224535063
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	7.6	Control (Path) 4 Temporal Ctx	0.0
AD 3 Hippo	0.0	AD 1 Occipital Ctx	0.0
AD 4 Hippo	7.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	8.2	AD 3 Occipital Ctx	0.0
AD 6 Hippo	29.5	AD 4 Occipital Ctx	0.0
Control 2 Hippo	14.2	AD 5 Occipital Ctx	24.5
Control 4 Hippo	4.1	AD 6 Occipital Ctx	21.2
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	1.7
AD 1 Temporal Ctx	0.0	Control 2 Occipital Ctx	18.2
AD 2 Temporal Ctx	8.7	Control 3 Occipital Ctx	9.2
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	0.0	Control (Path) 1 Occipital Ctx	0.0
AD 5 Inf Temporal Ctx	64.2	Control (Path) 2 Occipital Ctx	0.0
AD 5 Sup Temporal Ctx	11.0	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	100.0	Control (Path) 4 Occipital Ctx	7.1
AD 6 Sup Temporal Ctx	46.0	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	26.1
Control 2 Temporal Ctx	11.5	Control 3 Parietal Ctx	0.0
Control 3 Temporal Ctx	0.0	Control (Path) 1 Parietal Ctx	9.3
Control 4 Temporal Ctx	0.0	Control (Path) 2 Parietal Ctx	7.0
Control (Path) 1 Temporal Ctx	25.0	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	9.8	Control (Path) 4 Parietal Ctx	9.7

**Table NL. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4465, Run 222566842	Rel. Exp.(%) Ag4783, Run 217235304	Tissue Name	Rel. Exp.(%) Ag4465, Run 222566842	Rel. Exp.(%) Ag4783, Run 217235304
Adipose	4.6	12.5	Renal ca. TK-10	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	15.8	26.2
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell	0.0	0.0	Colon ca.* (SW480 met)	0.0	0.0



carcinoma SCC-4			SW620		
Testis Pool	13.2	12.9	Colon ca. HT29	0.0	0.0
Prostate ca. * (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	0.0	2.8	Colon ca. CaCo-2	0.0	0.0
Placenta	8.0	4.9	Colon cancer tissue	0.0	5.1
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.0	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	4.6	6.1
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	0.0	2.1
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	4.7	8.2
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	0.0	3.0
Ovary	0.0	2.4	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	4.0	0.0
Breast ca. MDA-MB-231	21.8	35.6	Lymph Node Pool	3.3	4.4
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	3.9	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	0.0	0.0	Spleen Pool	100.0	57.8
Breast Pool	0.0	4.3	Thymus Pool	14.4	11.0
Trachea	18.3	13.1	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.0	1.3	CNS cancer (glio/astro) U-118-MG	0.0	0.0
Fetal Lung	12.8	51.4	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	3.1	CNS cancer (astro) SNB-75	0.0	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	75.8	100.0	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. A549	0.0	2.9	Brain (Amygdala) Pool	15.4	6.2
Lung ca. NCI-H526	0.0	0.0	Brain (cerebellum)	5.3	13.5
Lung ca. NCI-H23	0.0	0.0	Brain (fetal)	0.0	9.9
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	26.4	30.6
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	14.5	2.1
Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	7.8	10.4
Liver	0.0	0.0	Brain (Thalamus) Pool	12.6	14.9
Fetal Liver	4.7	0.0	Brain (whole)	29.9	14.0

Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	21.6	15.0
Kidney Pool	0.0	0.0	Adrenal Gland	4.5	3.4
Fetal Kidney	0.0	0.0	Pituitary gland Pool	0.0	0.0
Renal ca. 786-0	0.0	0.0	Salivary Gland	10.5	0.0
Renal ca. A498	0.0	0.0	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	47.3	43.2

**Table NM. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4465, Run 191765065	Rel. Exp.(%) Ag4465, Run 195509496	Rel. Exp.(%) Ag4783, Run 209988936	Rel. Exp.(%) Ag4784, Run 209989013	Rel. Exp.(%) Ag5089, Run 223210710	Rel. Exp.(%) Ag5089, Run 229739338	Rel. Exp.(%) Ag6508, Run 271410131
Secondary Th1 act	1.5	1.8	1.4	2.1	0.0	4.6	1.9
Secondary Th2 act	1.0	1.6	1.9	1.1	0.0	3.0	3.1
Secondary Tr1 act	1.1	1.7	1.3	1.8	0.0	1.8	0.7
Secondary Th1 rest	0.9	0.6	0.5	0.6	0.0	0.0	0.0
Secondary Th2 rest	0.4	0.5	0.7	0.8	0.0	0.0	0.3
Secondary Tr1 rest	0.5	0.8	0.3	0.2	0.0	0.0	0.0
Primary Th1 act	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Primary Th2 act	0.2	0.3	0.0	0.1	0.0	0.0	0.0
Primary Tr1 act	0.2	0.1	0.1	0.2	0.0	0.0	0.0
Primary Th1 rest	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Primary Th2 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Tr1 rest	0.0	0.2	0.1	0.1	0.0	0.0	0.0
CD45RA CD4 lymphocyte act	0.1	0.1	81.2	0.0	0.0	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.1	0.1	0.0	0.0	0.0	0.0
CD8 lymphocyte act	0.0	0.0	0.1	0.2	0.0	0.0	0.0
Secondary CD8 lymphocyte rest	0.1	0.0	0.1	0.0	0.0	0.0	0.0
Secondary CD8 lymphocyte act	1.0	1.5	1.3	6.1	1.9	0.0	0.3
CD4 lymphocyte none	0.0	0.0	0.1	0.1	0.0	0.0	0.0

2ry Th1/Th2/Tr1_an ti-CD95 CH11	0.1	0.1	0.2	0.3	0.0	0.0	0.0
LAK cells rest	0.1	0.1	0.2	0.1	0.0	0.0	0.0
LAK cells IL-2	2.4	2.8	2.3	4.5	3.7	0.0	0.2
LAK cells IL-2+IL-12	0.2	0.2	0.1	0.2	0.0	0.0	0.0
LAK cells IL-2+IFN gamma	1.0	0.6	0.7	0.7	0.0	0.0	0.0
LAK cells IL-2+IL-18	1.0	0.8	0.9	0.9	0.0	0.0	0.0
LAK cells PMA/ionomycin	0.4	0.9	0.5	0.8	0.0	1.8	0.0
NK Cells IL-2 rest	3.4	4.4	3.9	3.3	3.1	19.5	12.4
Two Way MLR 3 day	0.3	0.5	0.3	0.4	0.0	0.0	0.0
Two Way MLR 5 day	0.2	0.2	0.3	0.3	0.0	0.0	0.0
Two Way MLR 7 day	0.1	0.2	0.1	0.3	0.0	0.0	0.0
PBMC rest	0.5	0.2	0.3	0.3	0.0	0.0	0.0
PBMC PWM	0.0	0.1	0.1	0.1	0.0	0.0	0.0
PBMC PHA-L	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Ramos (B cell) none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B lymphocytes PWM	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B lymphocytes CD40L and IL-4	0.1	0.0	0.0	0.0	0.0	0.0	0.0
EOL-1 dbcAMP	0.2	0.4	0.3	0.3	0.0	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Dendritic cells none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dendritic cells LPS	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocytes rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocytes LPS	0.0	0.0	0.1	0.2	0.0	0.0	0.0
Macrophages	0.0	0.0	0.0	0.0	0.0	0.0	0.0

rest							
Macrophages LPS	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC starved	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IFN gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC TNF alpha + IFN gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC TNF alpha + IL4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IL-11	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung Microvascular EC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Microvascular Dermal EC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bronchial epithelium TNFalpha + IL1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Small airway epithelium none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Small airway epithelium TNFalpha + IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coronary artery SMC rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Astrocytes rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Astrocytes TNFalpha + IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0

KU-812 (Basophil) rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
KU-812 (Basophil) PMA/ionomycin	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CCD1106 (Keratinocytes) none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CCD1106 (Keratinocytes) TNFalpha + IL- 1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Liver cirrhosis	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 IL-4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 IL-9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 IL-13	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 IFN gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HPAEC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HPAEC TNF alpha + IL-1 beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast TNF alpha + IL- 1 beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast IL-4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast IL-9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast IL-13	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast IFN gamma	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Dermal fibroblast CCD1070 rest	0.2	0.1	0.0	0.0	0.0	0.0	0.0
Dermal fibroblast CCD1070 TNF alpha	2.7	3.6	5.0	6.2	2.5	27.5	9.3
Dermal fibroblast CCD1070 IL-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

beta							
Dermal fibroblast IFN gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dermal fibroblast IL-4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dermal Fibroblasts rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Neutrophils TNFa+LPS	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Neutrophils rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Colon	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Thymus	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Kidney	0.0	0.0	0.0	0.1	0.0	0.0	0.0

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag4783 Two experiments with same probe and primer sets are in good agreement with highest expression of this gene in samples derived from ulcerative colitis patient (CT=32-33). In addition, low levels of expression of this gene seem to be restricted to emphysema, psoriasis and ulcerative colitis samples. Therefore, expression of this gene may be used to differentiate these samples from other samples in this panel. Furthermore, therapeutic modulation of the protein encoded by this gene may be useful in the treatment of emphysema, psoriasis and inflammatory bowel diseases including ulcerative colitis. A third run with this probe and primer set, run 211063353, shows low/undetectable levels of expression across all samples on this panel (CTs>35). (Data not shown.)

Ag5089/Ag6237/Ag6247 Expression is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). Please note that probes Ag6237 and Ag6247 are specific for CG112813-05 and CG112813-06 respectively.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4465 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Low expression of this gene in the brain suggests that may play a role in central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**General\_screening\_panel\_v1.4 Summary:** Ag4465/Ag4783 Two experiments with different probe and primer sets are in good agreement, with highest expression of the

CG112813-01 gene in spleen and a lung cancer SHP-77 cell line. Low levels of expression of this gene is also detected in breast cancer cell line. Therefore, expression of this gene may be used as marker for detection of lung and breast cancer. Furthermore, expression of this gene in spleen suggests that this gene may be involved in secondary immune  
5 responses. Therefore, antibodies or small molecule therapeutics that block the function of the protein encoded by this gene may be useful as anti-inflammatory therapeutics for the treatment of allergies, autoimmune diseases, and inflammatory diseases.

In addition, low levels of expression of this gene is also seen in pancreas. Therefore, therapeutic modulation of the protein encoded by this gene may be useful in the treatment  
10 of disease related to pancreas including obesity and diabetes.

**General\_screening\_panel\_v1.5 Summary:** Ag5089 Expression of the CG112813-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Ag6237 Expression of the CG112813-05 gene is low/undetectable (CTs > 35)  
15 across all of the samples on this panel (data not shown).

Ag6247 Expression of the CG112813-06 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**Panel 4.1D Summary:** Ag4465/Ag4783/Ag4784/Ag5089 Results from six experiments with four different probes and primer sets are in good agreement, with highest  
20 expression of the CG112813-01 gene in ionmycin/PMA treated eosinophils (CTs=24.6-34). Therefore, expression of this gene may be used to differentiate this sample from other samples used in the panel.

Eosinophils and the cytokines and the inflammatory mediators produced by them may contribute to the pathology of inflammatory bowel diseases (IBD) (1) and asthma (2).  
25 IBD including Crohn's disease and ulcerative colitis are strongly associated with infiltration of eosinophils. Eosinophils are the most prevalent granulocyte present in acute IBD (1). Eosinophil products such as IL-16 (3) and TGF alpha (4) appear to be involved in the inflammation and subsequent chronic changes associated with this disease.

The role of eosinophils in asthma has recently been brought into question by recent  
30 phase I clinical trials with anti-IL-5 Mabs SCH55700 and SB-240563 (reviewed in 3). IL-5 is important in the generation of eosinophils in the bone marrow and survival of eosinophils in the periphery. Thus, eosinophils remain an important cellular therapeutic target in the treatment of asthma.

The CG112813 gene codes for a protein belonging to killer cell immunoglobulin (Ig)-like receptors family (KIR; 5) on chromosome 19. KIR are MHC class I-binding immunoreceptors that can suppress activation of human NK cells (5, 6). NK cells have been shown to regulate inflammation and intervene in loss of self-tolerance (7). Therefore, the KIR protein encoded by this gene may also play a role in regulation of NK cells and thus, play a role in regulation of autoimmune diseases.

In addition, moderate to low levels of expression of this gene is also seen in TNF alpha treated dermal fibroblasts, IL-2 treated NK cells, cytokine treated LAK cells, activated secondary CD8 lymphocytes, and secondary Th1, Th2, and Tr1 cells. Since these cells, including eosinophils, play an important role in lung pathology, inflammatory bowel disease, and autoimmune disorders, including rheumatoid arthritis, antibody or small molecule therapies designed with the protein encoded by this gene may block or inhibit inflammation and tissue resulting from asthma, allergies, hypersensitivity reactions, inflammatory bowel disease, psoriasis, emphysema, viral infections and autoimmune diseases.

Ag6508 Results with one probe and primer set specific to CG112813-05 are in agreement with results presented above.

Highest expression in this experiment was seen in ionmycin/PMA treated eosinophils (CT=31).

Ag6237 Expression of the CG112813-05 gene is low/undetectable (CTs > 35) across all of the samples on this panel due to a probable probe or chemistry failure (data not shown).

Ag6247 Expression of the CG112813-06 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). Results from one experiment (Run 258176981) with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

#### References:

1. Jeziorska M, Haboubi N, Schofield P, Woolley DE. Distribution and activation of eosinophils in inflammatory bowel disease using an improved immunohistochemical technique. J Pathol 2001 Aug;194(4):484-92.

2. Seegert D, Rosenstiel P, Pfahler H, Pfefferkorn P, Nikolaus S, Schreiber S. Increased expression of IL-16 in inflammatory bowel disease. Gut 2001 Mar;48(3):326-32.

3. Grip O, Malm J, Veress B, Bjartell A, Lindgren S, Eggesten A. Increased presence of cells containing transforming growth factor alpha (TGF-alpha) in ulcerative colitis, both



during active inflammation and in remission. Eur J Gastroenterol Hepatol 2000 Jul;12(7):761-6

4. Trifilieff A, Fujitani Y, Coyle AJ, Kopf M, Bertrand C. IL-5 deficiency abolishes aspects of airway remodelling in a murine model of lung inflammation. Clin Exp Allergy 2001 Jun;31(6):934-42.

5. Ann. Rev. Immunol. 16:359-93. 1998. NK Cell Receptors. Lanier, Lewis.

6. Yusa S, Catina TL, Campbell KS. SHP-1- and Phosphotyrosine-Independent Inhibitory Signaling by a Killer Cell Ig-Like Receptor Cytoplasmic Domain in Human NK Cells. J Immunol 2002 May 15;168(10):5047-57, PMID: 11994457.

7. Flodstrom M, Shi FD, Sarvetnick N, Ljunggren HG. The natural killer cell - friend or foe in autoimmune disease? Scand J Immunol 2002 May;55(5):432-41, PMID: 11975754 .

**general oncology screening panel\_v\_2.4 Summary:** Ag4465 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

#### **O. CG112869-01: Pecanex like.**

Expression of gene CG112869-01 was assessed using the primer-probe set Ag6810, described in Table OA.

**Table OA. Probe Name Ag6810**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ggtacccagctgatgatcatt-3'	21	1588	223
Probe	TET-5' -cagaatgatgtccgcaacagcttcatt-3' -TAMRA	26	1615	224
Reverse	5' -agtgaagctcaagttaataaactggttaa-3'	28	1650	225

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6810 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.6 Summary:** Ag6810 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 4.1D Summary:** Ag6810 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

#### **P. CG113377-01: G1-Related zinc finger protein**

Expression of gene CG113377-01 was assessed using the primer-probe set Ag6802, described in Table PA. Results of the RTQ-PCR runs are shown in Tables PB, PC and PD.

**Table PA. Probe Name Ag6802**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gttgtgtttgtctccatctcctt-3'	23	632	226
Probe	TET-5'-attgtcctgatgatcatttcctctgc-3'-TAMRA	26	656	227
Reverse	5'-tatcgaaacctctggatgtaataaaa-3'	26	692	228

**Table PB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag6802, Run 278022718	Tissue Name	Rel. Exp.(%) Ag6802, Run 278022718
AD 1 Hippo	15.0	Control (Path) 3 Temporal Ctx	6.0
AD 2 Hippo	21.9	Control (Path) 4 Temporal Ctx	29.3
AD 3 Hippo	12.2	AD 1 Occipital Ctx	9.4
AD 4 Hippo	9.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	77.9	AD 3 Occipital Ctx	7.1
AD 6 Hippo	31.0	AD 4 Occipital Ctx	22.1
Control 2 Hippo	46.7	AD 5 Occipital Ctx	42.0
Control 4 Hippo	11.8	AD 6 Occipital Ctx	11.3
Control (Path) 3 Hippo	5.8	Control 1 Occipital Ctx	4.5
AD 1 Temporal Ctx	14.6	Control 2 Occipital Ctx	31.6
AD 2 Temporal Ctx	45.1	Control 3 Occipital Ctx	20.4
AD 3 Temporal Ctx	6.1	Control 4 Occipital Ctx	6.1
AD 4 Temporal Ctx	24.0	Control (Path) 1 Occipital Ctx	90.1
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.9
AD 5 Sup Temporal Ctx	85.9	Control (Path) 3 Occipital Ctx	2.0
AD 6 Inf Temporal Ctx	48.6	Control (Path) 4 Occipital Ctx	11.2
AD 6 Sup Temporal Ctx	42.3	Control 1 Parietal Ctx	8.7
Control 1 Temporal Ctx	6.3	Control 2 Parietal Ctx	32.5
Control 2 Temporal Ctx	44.4	Control 3 Parietal Ctx	27.2
Control 3 Temporal Ctx	15.4	Control (Path) 1 Parietal Ctx	81.2
Control 3 Temporal Ctx	14.4	Control (Path) 2 Parietal Ctx	26.1
Control (Path) 1 Temporal Ctx	99.3	Control (Path) 3 Parietal Ctx	3.8
Control (Path) 2 Temporal Ctx	50.0	Control (Path) 4 Parietal Ctx	36.6

**Table PC. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6802, Run 278017575	Tissue Name	Rel. Exp.(%) Ag6802, Run 278017575
Adipose	6.9	Renal ca. TK-10	0.1

Melanoma* Hs688(A).T	10.4	Bladder	2.5
Melanoma* Hs688(B).T	3.7	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	3.6	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	20.4	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.3	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	3.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.3	Colon ca. HCT-116	0.0
Prostate Pool	5.6	Colon ca. CaCo-2	0.0
Placenta	1.0	Colon cancer tissue	1.5
Uterus Pool	2.3	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	24.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	5.6	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	1.9	Colon Pool	18.6
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	14.1
Ovarian ca. IGROV-1	2.3	Stomach Pool	9.1
Ovarian ca. OVCAR-8	4.4	Bone Marrow Pool	5.7
Ovary	7.7	Fetal Heart	31.4
Breast ca. MCF-7	2.3	Heart Pool	11.7
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	19.1
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	11.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	2.9
Breast ca. MDA-N	3.1	Spleen Pool	7.4
Breast Pool	20.2	Thymus Pool	11.0
Trachea	2.6	CNS cancer (glio/astro) U87-MG	0.1
Lung	2.9	CNS cancer (glio/astro) U-118-MG	19.8
Fetal Lung	11.7	CNS cancer (neuro;met) SK-N-AS	100.0
Lung ca. NCI-N417	0.4	CNS cancer (astro) SF-539	11.1
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	6.5
Lung ca. NCI-H146	0.3	CNS cancer (glio) SNB-19	1.7
Lung ca. SHP-77	1.6	CNS cancer (glio) SF-295	0.7
Lung ca. A549	0.5	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	1.4	Brain (cerebellum)	19.6
Lung ca. NCI-H23	1.7	Brain (fetal)	25.5
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	4.8
Lung ca. HOP-62	3.0	Cerebral Cortex Pool	6.9
Lung ca. NCI-H522	2.3	Brain (Substantia nigra) Pool	3.9
Liver	0.1	Brain (Thalamus) Pool	9.3
Fetal Liver	1.1	Brain (whole)	8.8

Liver ca. HepG2	0.0	Spinal Cord Pool	3.8
Kidney Pool	22.4	Adrenal Gland	23.2
Fetal Kidney	13.0	Pituitary gland Pool	4.4
Renal ca. 786-0	0.0	Salivary Gland	19.8
Renal ca. A498	0.1	Thyroid (female)	2.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.1	Pancreas Pool	0.5

**Table PD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag6802, Run 278020684	Tissue Name	Rel. Exp.(%) Ag6802, Run 278020684
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.9
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.8
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.6
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	35.1	Coronary artery SMC rest	8.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	15.0
CD8 lymphocyte act	0.0	Astrocytes rest	7.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	1.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.7
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.5
LAK cells IL-2	0.0	Liver cirrhosis	10.2
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0

LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	2.2
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	14.8
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	19.6
PBMC PWM	0.0	Lung fibroblast IL-4	8.2
PBMC PHA-L	0.0	Lung fibroblast IL-9	5.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	5.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	8.4
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	21.0
B lymphocytes CD40L and IL-4	0.5	Dermal fibroblast CCD1070 TNF alpha	41.5
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	18.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	10.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	25.5
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	13.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	2.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	5.1
Macrophages rest	0.0	Lung	4.2
Macrophages LPS	0.0	Thymus	6.0
HUVEC none	0.5	Kidney	100.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6802 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.6 for discussion of utility of this gene in the central nervous system.

5 **General\_screening\_panel\_v1.6 Summary:** Ag6802 Highest expression of this gene in this panel is seen in a brain cancer cell line (CT=26.7). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of brain cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of brain cancer.

10 Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, fetal liver, and adult and fetal skeletal muscle and heart. This widespread expression among these tissues suggests that

this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag6802 This gene is most highly expressed in the kidney (CT=31.8). Thus, expression of this gene could be used to differentiate the kidney derived sample from other samples on this panel and as a marker of kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

#### Q. CG113730-01: NODAL PRECURSOR

Expression of gene CG113730-01 was assessed using the primer-probe set Ag4473, described in Table QA. Results of the RTQ-PCR runs are shown in Tables QB, QC, QD and QE.

**Table QA. Probe Name Ag4473**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - aagtcaactgtgtcgggaaggt - 3'	21	789	229
Probe	TET-5' - caagttccaggtggacttcaacctga - 3' - TAMRA	26	810	230
Reverse	5' - gttgtactgcttgggtagatg - 3'	22	855	231

**Table QB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4473, Run 224535202	Tissue Name	Rel. Exp.(%) Ag4473, Run 224535202
AD 1 Hippo	13.7	Control (Path) 3 Temporal Ctx	9.1
AD 2 Hippo	41.5	Control (Path) 4 Temporal Ctx	62.0
AD 3 Hippo	4.3	AD 1 Occipital Ctx	37.6
AD 4 Hippo	18.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	45.4	AD 3 Occipital Ctx	4.5
AD 6 Hippo	79.6	AD 4 Occipital Ctx	66.9
Control 2 Hippo	27.4	AD 5 Occipital Ctx	49.7

Control 4 Hippo	27.0	AD 6 Occipital Ctx	24.0
Control (Path) 3 Hippo	5.1	Control 1 Occipital Ctx	10.0
AD 1 Temporal Ctx	40.6	Control 2 Occipital Ctx	73.2
AD 2 Temporal Ctx	37.4	Control 3 Occipital Ctx	46.3
AD 3 Temporal Ctx	1.2	Control 4 Occipital Ctx	20.2
AD 4 Temporal Ctx	84.1	Control (Path) 1 Occipital Ctx	70.7
AD 5 Inf Temporal Ctx	59.5	Control (Path) 2 Occipital Ctx	17.6
AD 5 Sup Temporal Ctx	57.8	Control (Path) 3 Occipital Ctx	8.1
AD 6 Inf Temporal Ctx	79.6	Control (Path) 4 Occipital Ctx	12.9
AD 6 Sup Temporal Ctx	75.8	Control 1 Parietal Ctx	22.7
Control 1 Temporal Ctx	17.6	Control 2 Parietal Ctx	42.0
Control 2 Temporal Ctx	27.2	Control 3 Parietal Ctx	21.6
Control 3 Temporal Ctx	18.7	Control (Path) 1 Parietal Ctx	84.1
Control 3 Temporal Ctx	31.9	Control (Path) 2 Parietal Ctx	28.7
Control (Path) 1 Temporal Ctx	100.0	Control (Path) 3 Parietal Ctx	8.2
Control (Path) 2 Temporal Ctx	27.9	Control (Path) 4 Parietal Ctx	100.0

**Table QC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4473, Run 222655834	Tissue Name	Rel. Exp.(%) Ag4473, Run 222655834
Adipose	7.0	Renal ca. TK-10	10.7
Melanoma* Hs688(A).T	5.0	Bladder	19.5
Melanoma* Hs688(B).T	5.5	Gastric ca. (liver met.) NCI-N87	52.5
Melanoma* M14	2.7	Gastric ca. KATO III	8.7
Melanoma* LOXIMV1	0.0	Colon ca. SW-948	2.4
Melanoma* SK-MEL-5	2.4	Colon ca. SW480	37.1
Squamous cell carcinoma SCC-4	2.7	Colon ca.* (SW480 met) SW620	20.9
Testis Pool	9.2	Colon ca. HT29	4.4
Prostate ca.* (bone met) PC-3	14.2	Colon ca. HCT-116	15.3
Prostate Pool	5.7	Colon ca. CaCo-2	100.0
Placenta	8.9	Colon cancer tissue	7.8
Uterus Pool	11.7	Colon ca. SW1116	3.9
Ovarian ca. OVCAR-3	16.5	Colon ca. Colo-205	3.1
Ovarian ca. SK-OV-3	10.2	Colon ca. SW-48	20.0
Ovarian ca. OVCAR-4	2.6	Colon Pool	19.1
Ovarian ca. OVCAR-5	26.4	Small Intestine Pool	33.4
Ovarian ca. IGROV-1	6.0	Stomach Pool	13.9
Ovarian ca. OVCAR-8	6.0	Bone Marrow Pool	12.7

Ovary	9.7	Fetal Heart	13.6
Breast ca. MCF-7	5.8	Heart Pool	6.8
Breast ca. MDA-MB-231	24.0	Lymph Node Pool	27.7
Breast ca. BT 549	25.9	Fetal Skeletal Muscle	19.3
Breast ca. T47D	47.3	Skeletal Muscle Pool	27.5
Breast ca. MDA-N	2.7	Spleen Pool	2.8
Breast Pool	21.5	Thymus Pool	14.7
Trachea	5.4	CNS cancer (glio/astro) U87-MG	2.2
Lung	9.7	CNS cancer (glio/astro) U-118-MG	2.6
Fetal Lung	44.4	CNS cancer (neuro;met) SK-N-AS	48.3
Lung ca. NCI-N417	1.9	CNS cancer (astro) SF-539	5.6
Lung ca. LX-1	22.5	CNS cancer (astro) SNB-75	16.8
Lung ca. NCI-H146	16.4	CNS cancer (glio) SNB-19	4.7
Lung ca. SHP-77	18.3	CNS cancer (glio) SF-295	18.8
Lung ca. A549	17.3	Brain (Amygdala) Pool	5.4
Lung ca. NCI-H526	2.9	Brain (cerebellum)	27.2
Lung ca. NCI-H23	37.9	Brain (fetal)	54.0
Lung ca. NCI-H460	4.5	Brain (Hippocampus) Pool	5.0
Lung ca. HOP-62	4.5	Cerebral Cortex Pool	7.5
Lung ca. NCI-H522	47.3	Brain (Substantia nigra) Pool	12.8
Liver	0.0	Brain (Thalamus) Pool	6.8
Fetal Liver	3.9	Brain (whole)	13.5
Liver ca. HepG2	3.6	Spinal Cord Pool	7.1
Kidney Pool	60.7	Adrenal Gland	5.4
Fetal Kidney	17.1	Pituitary gland Pool	1.5
Renal ca. 786-0	11.0	Salivary Gland	0.6
Renal ca. A498	3.4	Thyroid (female)	1.1
Renal ca. ACHN	6.6	Pancreatic ca. CAPAN2	6.7
Renal ca. UO-31	2.3	Pancreas Pool	14.8

**Table QD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4473, Run 191882145	Tissue Name	Rel. Exp.(%) Ag4473, Run 191882145
Secondary Th1 act	14.7	HUVEC IL-1beta	13.7
Secondary Th2 act	3.4	HUVEC IFN gamma	7.6
Secondary Tr1 act	17.7	HUVEC TNF alpha + IFN gamma	6.7
Secondary Th1 rest	3.8	HUVEC TNF alpha + IL4	15.5
Secondary Th2 rest	30.8	HUVEC IL-11	4.1



Secondary Tr1 rest	21.5	Lung Microvascular EC none	15.6
Primary Th1 act	6.8	Lung Microvascular EC TNFalpha + IL-1beta	13.7
Primary Th2 act	35.4	Microvascular Dermal EC none	3.5
Primary Tr1 act	11.9	Microvascular Dermal EC TNFalpha + IL-1beta	12.9
Primary Th1 rest	11.9	Bronchial epithelium TNFalpha + IL1beta	16.4
Primary Th2 rest	7.0	Small airway epithelium none	6.4
Primary Tr1 rest	7.9	Small airway epithelium TNFalpha + IL-1beta	13.2
CD45RA CD4 lymphocyte act	8.4	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	30.6	Coronary artery SMC TNFalpha + IL-1beta	11.6
CD8 lymphocyte act	0.0	Astrocytes rest	14.0
Secondary CD8 lymphocyte rest	1.9	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	3.5	KU-812 (Basophil) rest	3.3
CD4 lymphocyte none	28.5	KU-812 (Basophil) PMA/ionomycin	14.4
2ry Th1/Th2/Tr1 _anti-CD95 CH11	20.9	CCD1106 (Keratinocytes) none	7.3
LAK cells rest	7.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.0
LAK cells IL-2	21.9	Liver cirrhosis	8.2
LAK cells IL-2+IL-12	7.1	NCI-H292 none	17.4
LAK cells IL-2+IFN gamma	7.2	NCI-H292 IL-4	17.6
LAK cells IL-2+ IL-18	11.4	NCI-H292 IL-9	40.3
LAK cells PMA/ionomycin	7.3	NCI-H292 IL-13	64.6
NK Cells IL-2 rest	54.0	NCI-H292 IFN gamma	29.7
Two Way MLR 3 day	24.7	HPAEC none	0.0
Two Way MLR 5 day	11.0	HPAEC TNF alpha + IL-1 beta	4.4
Two Way MLR 7 day	24.5	Lung fibroblast none	33.7
PBMC rest	13.8	Lung fibroblast TNF alpha + IL-1 beta	5.1
PBMC PWM	9.0	Lung fibroblast IL-4	7.4
PBMC PHA-L	0.0	Lung fibroblast IL-9	1.8
Ramos (B cell) none	16.4	Lung fibroblast IL-13	4.3
Ramos (B cell) ionomycin	11.9	Lung fibroblast IFN gamma	1.9
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	3.7
B lymphocytes CD40L and IL-4	23.8	Dermal fibroblast CCD1070 TNF alpha	20.0
EOL-1 dbcAMP	100.0	Dermal fibroblast CCD1070 IL-1 beta	15.8
EOL-1 dbcAMP PMA/ionomycin	29.5	Dermal fibroblast IFN gamma	4.7
Dendritic cells none	23.8	Dermal fibroblast IL-4	1.9

Dendritic cells LPS	3.8	Dermal Fibroblasts rest	7.3
Dendritic cells anti-CD40	13.6	Neutrophils TNFa+LPS	13.8
Monocytes rest	30.4	Neutrophils rest	25.7
Monocytes LPS	31.2	Colon	7.9
Macrophages rest	10.4	Lung	13.7
Macrophages LPS	0.0	Thymus	29.7
HUVEC none	1.8	Kidney	26.4
HUVEC starved	15.1		

**Table QE. general oncology screening panel\_v\_2.4**

Tissue Name	Rel. Exp.(%) Ag4473, Run 268672309	Tissue Name	Rel. Exp.(%) Ag4473, Run 268672309
Colon cancer 1	8.4	Bladder NAT 2	4.0
Colon NAT 1	3.0	Bladder NAT 3	0.0
Colon cancer 2	11.1	Bladder NAT 4	2.1
Colon NAT 2	10.8	Prostate adenocarcinoma 1	53.2
Colon cancer 3	23.0	Prostate adenocarcinoma 2	0.0
Colon NAT 3	26.1	Prostate adenocarcinoma 3	6.5
Colon malignant cancer 4	7.0	Prostate adenocarcinoma 4	10.9
Colon NAT 4	0.7	Prostate NAT 5	8.8
Lung cancer 1	17.6	Prostate adenocarcinoma 6	1.5
Lung NAT 1	2.5	Prostate adenocarcinoma 7	7.7
Lung cancer 2	32.3	Prostate adenocarcinoma 8	0.0
Lung NAT 2	12.9	Prostate adenocarcinoma 9	25.2
Squamous cell carcinoma 3	9.6	Prostate NAT 10	0.0
Lung NAT 3	2.4	Kidney cancer 1	27.0
Metastatic melanoma 1	28.3	Kidney NAT 1	8.9
Melanoma 2	4.2	Kidney cancer 2	42.3
Melanoma 3	0.0	Kidney NAT 2	24.8
Metastatic melanoma 4	100.0	Kidney cancer 3	65.1
Metastatic melanoma 5	70.2	Kidney NAT 3	8.1
Bladder cancer 1	2.9	Kidney cancer 4	7.1
Bladder NAT 1	0.0	Kidney NAT 4	3.5
Bladder cancer 2	4.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4473 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene at low levels in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

**General\_screening\_panel\_v1.4 Summary:** Ag4473 Highest expression of this gene is seen in a colon cancer cell line (CT=30.9). This gene is widely expressed among the cancer cell lines on this panel, with moderate to low expression seen in brain, colon, gastric, lung, breast, and ovarian cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at low but significant levels in adipose, pancreas, and adult and fetal skeletal muscle, heart, and liver. This expression suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the thalamus, substantia nigra, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag4473 Highest expression is seen in eosinophils (CT=32.5). Low but significant expression is also seen in many other cell types of significance in the immune response in health and disease. These cells include members of the T-cell and B-cell family. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

**general oncology screening panel\_v\_2.4 Summary:** Ag4473 This gene is widely expressed in this panel, with highest expression in metastatic melanoma cancer (CT=32.5). In addition, this gene is moderately expressed in prostate cancer. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of prostate and melanoma cancer.

**R. CG115187-01 and CG115187-02: Novel human transmembrane protein**

Expression of gene CG115187-01 and variant CG115187-02 was assessed using the primer-probe sets Ag4480 and Ag5887, described in Tables RA and RB. Results of the RTQ-PCR runs are shown in Tables RC, RD, RE, RF, RG, RH, RI, RJ and RK.

5

**Table RA. Probe Name Ag4480**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gccagacacattgatctgaaac-3'	22	347	232
Probe	TET-5'-ctaaccggtgccattatgttcccaag-3'-TAMRA	26	378	233
Reverse	5'-cttatcacctcctcagcttct-3'	22	414	234

**Table RB. Probe Name Ag5887**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttgcttattgtgtcgtgttaa-3'	22	110	235
Probe	TET-5'-ctgcaaccaccaggaccagaatgt-3'-TAMRA	25	150	236
Reverse	5'-cggggtaataacctcctacagt-3'	22	176	237

**Table RC. AI\_comprehensive panel\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4480, Run 228397223	Tissue Name	Rel. Exp.(%) Ag4480, Run 228397223
I10967 COPD-F	2.0	I12427 Match Control Psoriasis-F	32.5
I10980 COPD-F	30.1	I12418 Psoriasis-M	2.0
I10968 COPD-M	1.9	I12723 Match Control Psoriasis-M	1.4
I10977 COPD-M	26.8	I12419 Psoriasis-M	3.5
I10989 Emphysema-F	3.8	I12424 Match Control Psoriasis-M	4.0
I10992 Emphysema-F	1.3	I12420 Psoriasis-M	9.4
I10993 Emphysema-F	3.2	I12425 Match Control Psoriasis-M	22.4
I10994 Emphysema-F	2.0	I04689 (MF) OA Bone-Backus	14.9
I10995 Emphysema-F	4.5	I04690 (MF) Adj "Normal" Bone-Backus	43.2
I10996 Emphysema-F	0.0	I04691 (MF) OA Synovium-Backus	29.1
I10997 Asthma-M	21.6	I04692 (BA) OA Cartilage-Backus	0.0
I11001 Asthma-F	0.5	I04694 (BA) OA Bone-Backus	1.1
I11002 Asthma-F	0.9	I04695 (BA) Adj "Normal" Bone-Backus	25.0
I11003 Atopic Asthma-F	2.3	I04696 (BA) OA Synovium-Backus	7.2
I11004 Atopic Asthma-F	1.7	I04700 (SS) OA Bone-Backus	31.2
I11005 Atopic Asthma-F	1.5	I04701 (SS) Adj "Normal" Bone-Backus	18.4

111006 Atopic Asthma-F	0.7	104702 (SS) OA Synovium-Backus	19.2
111417 Allergy-M	2.6	117093 OA Cartilage Rep7	8.4
112347 Allergy-M	0.0	112672 OA Bone5	9.9
112349 Normal Lung-F	0.0	112673 OA Synovium5	6.5
112357 Normal Lung-F	92.0	112674 OA Synovial Fluid cells5	4.6
112354 Normal Lung-M	100.0	117100 OA Cartilage Rep14	2.0
112374 Crohns-F	1.1	112756 OA Bone9	0.8
112389 Match Control Crohns-F	11.3	112757 OA Synovium9	1.4
112375 Crohns-F	1.8	112758 OA Synovial Fluid Cells9	1.6
112732 Match Control Crohns-F	4.1	117125 RA Cartilage Rep2	3.0
112725 Crohns-M	0.0	113492 Bone2 RA	8.7
112387 Match Control Crohns-M	3.8	113493 Synovium2 RA	4.9
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	4.3
112390 Match Control Crohns-M	6.7	113499 Cartilage4 RA	5.9
112726 Crohns-M	4.4	113500 Bone4 RA	6.8
112731 Match Control Crohns-M	19.8	113501 Synovium4 RA	3.5
112380 Ulcer Col-F	5.5	113502 Syn Fluid Cells4 RA	2.2
112734 Match Control Ulcer Col-F	13.7	113495 Cartilage3 RA	6.6
112384 Ulcer Col-F	4.1	113496 Bone3 RA	7.9
112737 Match Control Ulcer Col-F	6.7	113497 Synovium3 RA	4.9
112386 Ulcer Col-F	8.8	113498 Syn Fluid Cells3 RA	13.2
112738 Match Control Ulcer Col-F	1.4	117106 Normal Cartilage Rep20	2.2
112381 Ulcer Col-M	2.9	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	0.0	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	16.5	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	2.7	117107 Normal Cartilage Rep22	2.4
112383 Ulcer Col-M	2.9	113667 Bone4 Normal	6.0
112736 Match Control Ulcer Col-M	20.2	113668 Synovium4 Normal	8.3
112423 Psoriasis-F	3.8	113669 Syn Fluid Cells4 Normal	5.1

**Table RD. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4480, Run 224535658	Tissue Name	Rel. Exp.(%) Ag4480, Run 224535658
AD 1 Hippo	31.2	Control (Path) 3 Temporal Ctx	8.0
AD 2 Hippo	84.7	Control (Path) 4 Temporal Ctx	39.2
AD 3 Hippo	16.0	AD 1 Occipital Ctx	29.1
AD 4 Hippo	12.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	36.6	AD 3 Occipital Ctx	15.9

AD 6 Hippo	99.3	AD 4 Occipital Ctx	29.9
Control 2 Hippo	49.3	AD 5 Occipital Ctx	24.3
Control 4 Hippo	57.8	AD 6 Occipital Ctx	33.0
Control (Path) 3 Hippo	8.9	Control 1 Occipital Ctx	4.0
AD 1 Temporal Ctx	44.8	Control 2 Occipital Ctx	33.2
AD 2 Temporal Ctx	46.3	Control 3 Occipital Ctx	16.0
AD 3 Temporal Ctx	12.2	Control 4 Occipital Ctx	32.3
AD 4 Temporal Ctx	32.3	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	83.5	Control (Path) 2 Occipital Ctx	13.1
AD 5 Sup Temporal Ctx	81.2	Control (Path) 3 Occipital Ctx	2.5
AD 6 Inf Temporal Ctx	74.7	Control (Path) 4 Occipital Ctx	17.9
AD 6 Sup Temporal Ctx	92.0	Control 1 Parietal Ctx	7.4
Control 1 Temporal Ctx	6.0	Control 2 Parietal Ctx	61.6
Control 2 Temporal Ctx	50.0	Control 3 Parietal Ctx	19.1
Control 3 Temporal Ctx	18.4	Control (Path) 1 Parietal Ctx	36.9
Control 4 Temporal Ctx	21.6	Control (Path) 2 Parietal Ctx	28.3
Control (Path) 1 Temporal Ctx	52.1	Control (Path) 3 Parietal Ctx	4.7
Control (Path) 2 Temporal Ctx	44.1	Control (Path) 4 Parietal Ctx	37.4

**Table RE. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4480, Run 222655888	Tissue Name	Rel. Exp.(%) Ag4480, Run 222655888
Adipose	2.7	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.9
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.4
Melanoma* M14	7.9	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	100.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	6.3	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.2	Colon ca. HT29	0.0
Prostate ca. * (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.4	Colon ca. CaCo-2	0.1
Placenta	0.0	Colon cancer tissue	0.9
Uterus Pool	1.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	3.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.4
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	3.5

Ovarian ca. IGROV-1	1.7	Stomach Pool	0.5
Ovarian ca. OVCAR-8	5.2	Bone Marrow Pool	2.7
Ovary	1.1	Fetal Heart	0.2
Breast ca. MCF-7	0.0	Heart Pool	0.5
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.6
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	2.2
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.5
Breast ca. MDA-N	0.0	Spleen Pool	0.3
Breast Pool	0.6	Thymus Pool	1.2
Trachea	3.0	CNS cancer (glio/astro) U87-MG	11.0
Lung	6.9	CNS cancer (glio/astro) U-118-MG	4.9
Fetal Lung	6.0	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	1.5	CNS cancer (astro) SF-539	16.5
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	3.3
Lung ca. NCI-H146	0.2	CNS cancer (glio) SNB-19	1.6
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	6.9
Lung ca. A549	0.0	Brain (Amygdala) Pool	2.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	8.4
Lung ca. NCI-H23	2.7	Brain (fetal)	1.7
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.6
Lung ca. HOP-62	1.5	Cerebral Cortex Pool	2.2
Lung ca. NCI-H522	0.2	Brain (Substantia nigra) Pool	2.2
Liver	0.0	Brain (Thalamus) Pool	5.2
Fetal Liver	0.2	Brain (whole)	3.4
Liver ca. HepG2	0.0	Spinal Cord Pool	13.2
Kidney Pool	1.2	Adrenal Gland	1.2
Fetal Kidney	0.5	Pituitary gland Pool	2.3
Renal ca. 786-0	0.0	Salivary Gland	0.2
Renal ca. A498	0.0	Thyroid (female)	0.2
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.6

**Table RF. General\_screening\_panel\_v1.5**

Tissue Name	Rel. Exp.(%) Ag4480, Run 228714893	Tissue Name	Rel. Exp.(%) Ag4480, Run 228714893
Adipose	2.5	Renal ca. TK-10	57.8
Melanoma* Hs688(A).T	34.6	Bladder	15.2
Melanoma* Hs688(B).T	27.2	Gastric ca. (liver met.) NCI-N87	72.2

Melanoma* M14	28.9	Gastric ca. KATO III	79.0
Melanoma* LOXIMVI	6.9	Colon ca. SW-948	12.3
Melanoma* SK-MEL-5	21.9	Colon ca. SW480	47.3
Squamous cell carcinoma SCC-4	8.4	Colon ca.* (SW480 met) SW620	13.4
Testis Pool	5.6	Colon ca. HT29	25.3
Prostate ca.* (bone met) PC-3	47.0	Colon ca. HCT-116	41.8
Prostate Pool	8.3	Colon ca. CaCo-2	24.3
Placenta	11.2	Colon cancer tissue	19.5
Uterus Pool	1.1	Colon ca. SW1116	5.4
Ovarian ca. OVCAR-3	25.9	Colon ca. Colo-205	20.7
Ovarian ca. SK-OV-3	39.5	Colon ca. SW-48	13.9
Ovarian ca. OVCAR-4	23.8	Colon Pool	10.2
Ovarian ca. OVCAR-5	49.0	Small Intestine Pool	5.7
Ovarian ca. IGROV-1	21.6	Stomach Pool	7.9
Ovarian ca. OVCAR-8	27.5	Bone Marrow Pool	2.6
Ovary	9.3	Fetal Heart	3.4
Breast ca. MCF-7	43.8	Heart Pool	5.1
Breast ca. MDA-MB-231	28.5	Lymph Node Pool	9.7
Breast ca. BT 549	71.7	Fetal Skeletal Muscle	4.8
Breast ca. T47D	8.1	Skeletal Muscle Pool	23.5
Breast ca. MDA-N	4.5	Spleen Pool	6.0
Breast Pool	9.3	Thymus Pool	6.6
Trachea	11.0	CNS cancer (glio/astro) U87-MG	73.2
Lung	3.1	CNS cancer (glio/astro) U-118-MG	48.6
Fetal Lung	15.2	CNS cancer (neuro;met) SK-N-AS	12.8
Lung ca. NCI-N417	5.5	CNS cancer (astro) SF-539	23.0
Lung ca. LX-1	37.4	CNS cancer (astro) SNB-75	100.0
Lung ca. NCI-H146	1.6	CNS cancer (glio) SNB-19	24.8
Lung ca. SHP-77	11.0	CNS cancer (glio) SF-295	66.0
Lung ca. A549	20.2	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	3.1	Brain (cerebellum)	10.0
Lung ca. NCI-H23	24.3	Brain (fetal)	2.1
Lung ca. NCI-H460	12.8	Brain (Hippocampus) Pool	6.0
Lung ca. HOP-62	34.4	Cerebral Cortex Pool	5.5
Lung ca. NCI-H522	41.5	Brain (Substantia nigra) Pool	5.6
Liver	27.5	Brain (Thalamus) Pool	6.7
Fetal Liver	80.7	Brain (whole)	7.2
Liver ca. HepG2	68.8	Spinal Cord Pool	6.0
Kidney Pool	12.4	Adrenal Gland	27.4



Fetal Kidney	14.7	Pituitary gland Pool	3.1
Renal ca. 786-0	40.9	Salivary Gland	9.8
Renal ca. A498	7.3	Thyroid (female)	13.2
Renal ca. ACHN	23.7	Pancreatic ca. CAPAN2	38.4
Renal ca. UO-31	62.0	Pancreas Pool	14.4

**Table RG. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag4480, Run 278391334	Tissue Name	Rel. Exp.(%) Ag4480, Run 278391334
Adipose	5.8	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	1.4
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.0
Melanoma* M14	10.7	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	100.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	4.8	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.9	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.7	Colon ca. CaCo-2	0.3
Placenta	0.1	Colon cancer tissue	1.8
Uterus Pool	2.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	5.2	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.5
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	5.4
Ovarian ca. IGROV-1	1.9	Stomach Pool	1.2
Ovarian ca. OVCAR-8	14.8	Bone Marrow Pool	4.1
Ovary	1.8	Fetal Heart	0.3
Breast ca. MCF-7	0.0	Heart Pool	0.9
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	1.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	4.9
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.4
Breast ca. MDA-N	0.0	Spleen Pool	0.8
Breast Pool	0.8	Thymus Pool	2.2
Trachea	3.2	CNS cancer (glio/astro) U87-MG	11.5
Lung	8.2	CNS cancer (glio/astro) U-118-MG	7.2
Fetal Lung	8.2	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	3.7	CNS cancer (astro) SF-539	20.9

Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	6.0
Lung ca. NCI-H146	0.2	CNS cancer (glio) SNB-19	1.9
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	10.7
Lung ca. A549	0.0	Brain (Amygdala) Pool	6.9
Lung ca. NCI-H526	0.0	Brain (cerebellum)	25.2
Lung ca. NCI-H23	2.7	Brain (fetal)	10.5
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	10.3
Lung ca. HOP-62	2.0	Cerebral Cortex Pool	5.0
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	3.3
Liver	0.1	Brain (Thalamus) Pool	9.3
Fetal Liver	0.2	Brain (whole)	5.5
Liver ca. HepG2	0.0	Spinal Cord Pool	21.2
Kidney Pool	5.3	Adrenal Gland	2.7
Fetal Kidney	0.5	Pituitary gland Pool	3.5
Renal ca. 786-0	0.0	Salivary Gland	0.5
Renal ca. A498	0.0	Thyroid (female)	0.4
Renal ca. ACHN	0.2	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.2	Pancreas Pool	0.1

**Table RH. HASS Panel v1.0**

Tissue Name	Rel. Exp.(%) Ag4480, Run 264749843	Tissue Name	Rel. Exp.(%) Ag4480, Run 264749843
MCF-7 C1	0.0	U87-MG F1 (B)	12.7
MCF-7 C2	0.0	U87-MG F2	4.3
MCF-7 C3	0.0	U87-MG F3	8.2
MCF-7 C4	0.0	U87-MG F4	13.4
MCF-7 C5	0.0	U87-MG F5	44.4
MCF-7 C6	0.0	U87-MG F6	69.7
MCF-7 C7	0.0	U87-MG F7	55.1
MCF-7 C9	0.0	U87-MG F8	50.3
MCF-7 C10	0.0	U87-MG F9	32.5
MCF-7 C11	0.0	U87-MG F10	62.4
MCF-7 C12	0.1	U87-MG F11	66.9
MCF-7 C13	0.0	U87-MG F12	45.4
MCF-7 C15	0.0	U87-MG F13	64.2
MCF-7 C16	0.0	U87-MG F14	54.3
MCF-7 C17	0.0	U87-MG F15	55.5
T24 D1	0.0	U87-MG F16	31.9
T24 D2	0.0	U87-MG F17	54.3

T24 D3	0.0	LnCAP A1	0.0
T24 D4	0.0	LnCAP A2	0.0
T24 D5	0.0	LnCAP A3	0.0
T24 D6	0.0	LnCAP A4	0.0
T24 D7	0.0	LnCAP A5	0.0
T24 D9	0.0	LnCAP A6	0.0
T24 D10	0.0	LnCAP A7	0.0
T24 D11	0.0	LnCAP A8	0.0
T24 D12	0.0	LnCAP A9	0.0
T24 D13	0.0	LnCAP A10	0.0
T24 D15	0.0	LnCAP A11	0.0
T24 D16	0.0	LnCAP A12	0.0
T24 D17	0.0	LnCAP A13	0.0
CAPaB B1	0.1	LnCAP A14	0.0
CAPaB B2	0.0	LnCAP A15	0.0
CAPaB B3	0.0	LnCAP A16	0.0
CAPaB B4	0.0	LnCAP A17	0.0
CAPaB B5	0.0	Primary Astrocytes	4.4
CAPaB B6	0.0	Primary Renal Proximal Tubule Epithelial cell A2	0.7
CAPaB B7	0.0	Primary melanocytes A5	3.4
CAPaB B8	0.0	126443 - 341 medullo	0.0
CAPaB B9	0.0	126444 - 487 medullo	0.0
CAPaB B10	0.0	126445 - 425 medullo	0.4
CAPaB B11	0.0	126446 - 690 medullo	0.7
CAPaB B12	0.0	126447 - 54 adult glioma	4.2
CAPaB B13	0.1	126448 - 245 adult glioma	1.6
CAPaB B14	0.0	126449 - 317 adult glioma	19.3
CAPaB B15	0.0	126450 - 212 glioma	100.0
CAPaB B16	0.0	126451 - 456 glioma	42.0
CAPaB B17	0.0		

**Table RI. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4480, Run 193613004	Rel. Exp.(%) Ag4480, Run 229739282	Tissue Name	Rel. Exp.(%) Ag4480, Run 193613004	Rel. Exp.(%) Ag4480, Run 229739282
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha +	0.0	0.0

			IFN gamma		
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	0.0
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	1.9
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.6	0.0
CD45RO CD4 lymphocyte act	1.4	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	9.3	6.9
Secondary CD8 lymphocyte rest	0.7	0.0	Astrocytes TNFalpha + IL-1beta	2.3	1.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.2	0.0	CCD1106 (Keratinocytes) none	5.5	1.5
LAK cells rest	11.8	5.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.5	0.0
LAK cells IL-2	14.2	7.0	Liver cirrhosis	4.1	3.3
LAK cells IL-2+IL-12	0.0	2.2	NCI-H292 none	1.7	1.6
LAK cells IL-2+IFN gamma	0.6	0.7	NCI-H292 IL-4	2.4	1.6
LAK cells IL-2+ IL-18	0.0	0.0	NCI-H292 IL-9	3.4	1.7
LAK cells PMA/ionomycin	9.6	8.3	NCI-H292 IL-13	0.8	1.0
NK Cells IL-2 rest	20.7	24.3	NCI-H292 IFN gamma	1.0	0.6
Two Way MLR 3 day	17.3	8.2	HPAEC none	0.0	0.0
Two Way MLR 5 day	5.3	0.0	HPAEC TNF alpha + IL-	0.0	0.0

			1 beta		
Two Way MLR 7 day	0.4	0.0	Lung fibroblast none	5.4	2.3
PBMC rest	5.7	0.8	Lung fibroblast TNF alpha + IL-1 beta	2.1	4.1
PBMC PWM	2.4	0.8	Lung fibroblast IL-4	2.3	1.6
PBMC PHA-L	2.4	1.6	Lung fibroblast IL-9	3.0	1.3
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.9	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IFN gamma	0.7	2.0
B lymphocytes PWM	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
B lymphocytes CD40L and IL-4	0.5	1.2	Dermal fibroblast CCD1070 TNF alpha	0.0	0.5
EOL-1 dbcAMP	3.1	1.3	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	2.0	0.0	Dermal fibroblast IFN gamma	0.7	0.0
Dendritic cells none	20.6	10.4	Dermal fibroblast IL-4	1.4	0.0
Dendritic cells LPS	22.2	17.4	Dermal Fibroblasts rest	1.4	2.0
Dendritic cells anti-CD40	48.3	24.7	Neutrophils TNFa+LPS	0.0	0.0
Monocytes rest	5.0	2.9	Neutrophils rest	0.0	0.0
Monocytes LPS	100.0	100.0	Colon	4.2	0.0
Macrophages rest	74.7	30.1	Lung	13.7	1.6
Macrophages LPS	42.6	6.5	Thymus	32.3	4.7
HUVEC none	0.0	0.0	Kidney	4.9	6.7
HUVEC starved	0.0	0.0			

Table RJ. Panel 5D

Tissue Name	Rel. Exp.(%) Ag5887, Run 258657708	Tissue Name	Rel. Exp.(%) Ag5887, Run 258657708
97457_Patient-02go_adipose	60.7	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	11.2	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	10.9	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.9
97481_Patient-08sk_skeletal muscle	7.1	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	9.7	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.9	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.8
97486_Patient-09sk_skeletal	0.0	94743_Donor 3 U - B_Mesenchymal Stem	0.0

muscle		Cells	
97487_Patient-09ut_uterus	3.6	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	2.7
97492_Patient-10ut_uterus	9.5	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.6	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	100.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	18.8	94735_Donor 3 AD - C_adipose	0.8
97497_Patient-11ut_uterus	24.0	77138_Liver_HepG2untreated	0.9
97498_Patient-11pl_placenta	1.8	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	69.7	81735_Small Intestine	7.6
97501_Patient-12sk_skeletal muscle	24.7	72409_Kidney_Proximal Convoluted Tubule	3.0
97502_Patient-12ut_uterus	5.3	82685_Small intestine_Duodenum	1.4
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	1.0	72410_Kidney_HRCE	14.7
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	6.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.8	73139_Uterus_Uterine smooth muscle cells	0.0

**Table RK. general oncology screening panel\_v\_2.4**

Tissue Name	Rel. Exp.(%) Ag4480, Run 260280485	Tissue Name	Rel. Exp.(%) Ag4480, Run 260280485
Colon cancer 1	1.5	Bladder cancer NAT 2	0.2
Colon cancer NAT 1	0.9	Bladder cancer NAT 3	0.0
Colon cancer 2	0.5	Bladder cancer NAT 4	2.3
Colon cancer NAT 2	1.9	Prostate adenocarcinoma 1	1.5
Colon cancer 3	2.4	Prostate adenocarcinoma 2	0.2
Colon cancer NAT 3	9.2	Prostate adenocarcinoma 3	0.7
Colon malignant cancer 4	7.0	Prostate adenocarcinoma 4	1.1
Colon normal adjacent tissue 4	3.2	Prostate cancer NAT 5	0.4
Lung cancer 1	2.5	Prostate adenocarcinoma 6	0.6
Lung NAT 1	0.6	Prostate adenocarcinoma 7	0.5
Lung cancer 2	3.5	Prostate adenocarcinoma 8	0.2
Lung NAT 2	0.7	Prostate adenocarcinoma 9	1.5
Squamous cell carcinoma 3	9.9	Prostate cancer NAT 10	0.3
Lung NAT 3	0.3	Kidney cancer 1	5.2

metastatic melanoma 1	68.3	KidneyNAT 1	2.7
Melanoma 2	0.7	Kidney cancer 2	7.4
Melanoma 3	1.3	Kidney NAT 2	6.0
metastatic melanoma 4	100.0	Kidney cancer 3	1.2
metastatic melanoma 5	74.2	Kidney NAT 3	2.2
Bladder cancer 1	0.1	Kidney cancer 4	8.7
Bladder cancer NAT 1	0.0	Kidney NAT 4	0.5
Bladder cancer 2	0.9		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag4480 This gene is widely expressed at low levels in many samples on this panel, with highest expression in normal lung (CT=31.5). Please see Panel 4.1D for discussion of utility of this gene in autoimmune disease.

5 **CNS\_neurodegeneration\_v1.0 Summary:** Ag4480 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

10 **General\_screening\_panel\_v1.4 Summary:** Ag4480 Highest expression of this gene in this panel is seen in a melanoma cell line (CT=27), with moderate levels of expression seen in brain cancer cell lines. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of melanoma. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma.

15 Among tissues with metabolic function, this gene is expressed at low but significant levels in pituitary, adipose, adrenal gland, pancreas, heart and adult and fetal skeletal muscle. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases,  
20 such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease,  
25 schizophrenia, multiple sclerosis, stroke and epilepsy.

**General\_screening\_panel\_v1.5 Summary:** Ag4480 Highest expression of this gene is seen in a brain cancer cell line (CT=30). This gene is widely expressed in this

panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

5        Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that  
10        dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

      This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease,  
15        schizophrenia, multiple sclerosis, stroke and epilepsy.

      Ag5887 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**General\_screening\_panel\_v1.6 Summary:** Ag4480 Highest expression of this gene in this panel is seen in a melanoma cell line (CT=28), with moderate levels of  
20        expression seen in brain and ovarian cancer cell lines. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of melanoma. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma.

      Among tissues with metabolic function, this gene is expressed at low but significant  
25        levels in pituitary, adipose, adrenal gland, heart and fetal skeletal muscle. This expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

      This gene is also expressed at moderate to low levels in the CNS, including the  
30        hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.



**HASS Panel v1.0 Summary:** Ag4480 This gene is expressed at a moderate level in the U87-MG cell line and the glioma samples on this panel suggesting it's role in brain cancer. The highest expression is seen in a glioma sample (CT= 28.55). Serum starvation induces expression of this gene in U87 cells suggesting that it may be used as a marker for areas of brain tumours that have poor vascularization.

**Panel 4.1D Summary:** Ag4480 Two experiments with the same probe and primer set produce results that are in reasonable agreement, with highest expression seen in LPS activated monocytes (CTs=32). In contrast, expression is undetectable in resting monocytes (CTs=36-37). Lower but substantial levels of expression are found in resting and activated dendritic cells and macrophages. Based on the expression pattern of this transcript, this gene product may be involved in monocyte activation and differentiation. Therefore, antibodies against the protein encoded by this gene may reduce or inhibit inflammation due to monocyte activation or differentiation and be important in the treatment of diseases such as asthma and arthritis. [Anrei Chapoval - GPDP]

**Panel 5D Summary:** Ag5887 Highest expression of this gene is seen in adipose (CT=31.3), with low but significant levels of expression detected in a cluster of samples derived from adipose and skeletal muscle. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

**general oncology screening panel\_v\_2.4 Summary:** Ag4480 Highest expression of this gene is seen in a melanoma sample (CT=29.7). Moderate levels of expression are also seen in a cluster of melanoma derived samples. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of melanoma. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma.

## **S. CG115187-03: transmembrane protein**

Expression of full length physical clone CG115187-03 was assessed using the primer-probe sets Ag5929 and Ag5887, described in Tables SA and SB. Results of the RTQ-PCR runs are shown in Tables SC, SD and SE.

**Table SA. Probe Name Ag5929**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cagaaggaagctgaggaggtta-3'	22	406	238
Probe	TET-5'-tcacctcagccgtgaattctgcaca-3'-TAMRA	25	435	239
Reverse	5'-gtcacagcagaagcaggtggtt-3'	21	475	240

**Table SB. Probe Name Ag5887**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - ttgcttattgtgtccgtgttaa - 3'	22	106	241
Probe	TET- 5' - ctgcaaccaccaggaccagaatgt - 3' - TAMRA	25	146	242
Reverse	5' - cggggtaataacctcctacagt - 3'	22	172	243

**Table SC. AI\_comprehensive panel\_v1.0**

Tissue Name	Rel. Exp.(%) Ag5929, Run 247834523	Tissue Name	Rel. Exp.(%) Ag5929, Run 247834523
110967 COPD-F	2.5	112427 Match Control Psoriasis-F	27.4
110980 COPD-F	30.8	112418 Psoriasis-M	1.5
110968 COPD-M	1.5	112723 Match Control Psoriasis-M	1.0
110977 COPD-M	28.3	112419 Psoriasis-M	4.8
110989 Emphysema-F	4.8	112424 Match Control Psoriasis-M	3.6
110992 Emphysema-F	3.3	112420 Psoriasis-M	5.3
110993 Emphysema-F	4.1	112425 Match Control Psoriasis-M	31.6
110994 Emphysema-F	1.4	104689 (MF) OA Bone-Backus	16.0
110995 Emphysema-F	5.0	104690 (MF) Adj "Normal" Bone-Backus	39.0
110996 Emphysema-F	0.2	104691 (MF) OA Synovium-Backus	15.8
110997 Asthma-M	21.3	104692 (BA) OA Cartilage-Backus	0.3
111001 Asthma-F	0.4	104694 (BA) OA Bone-Backus	0.5
111002 Asthma-F	2.6	104695 (BA) Adj "Normal" Bone-Backus	22.8
111003 Atopic Asthma-F	3.0	104696 (BA) OA Synovium-Backus	3.7
111004 Atopic Asthma-F	1.3	104700 (SS) OA Bone-Backus	28.7
111005 Atopic Asthma-F	0.5	104701 (SS) Adj "Normal" Bone-Backus	26.2
111006 Atopic Asthma-F	0.0	104702 (SS) OA Synovium-Backus	19.9
111417 Allergy-M	1.9	117093 OA Cartilage Rep7	5.9
112347 Allergy-M	0.1	112672 OA Bone5	17.8
112349 Normal Lung-F	0.0	112673 OA Synovium5	10.4
112357 Normal Lung-F	100.0	112674 OA Synovial Fluid cells5	3.7
112354 Normal Lung-M	100.0	117100 OA Cartilage Rep14	2.6
112374 Crohns-F	1.3	112756 OA Bone9	0.8
112389 Match Control Crohns-F	13.8	112757 OA Synovium9	1.2
112375 Crohns-F	2.3	112758 OA Synovial Fluid Cells9	0.9
112732 Match Control Crohns-F	2.7	117125 RA Cartilage Rep2	1.8
112725 Crohns-M	0.0	113492 Bone2 RA	7.2

112387 Match Control Crohns-M	2.0	113493 Synovium2 RA	6.7
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	7.5
112390 Match Control Crohns-M	7.6	113499 Cartilage4 RA	5.7
112726 Crohns-M	4.0	113500 Bone4 RA	5.8
112731 Match Control Crohns-M	12.2	113501 Synovium4 RA	4.5
112380 Ulcer Col-F	3.6	113502 Syn Fluid Cells4 RA	3.4
112734 Match Control Ulcer Col-F	11.2	113495 Cartilage3 RA	9.7
112384 Ulcer Col-F	5.8	113496 Bone3 RA	10.4
112737 Match Control Ulcer Col-F	6.1	113497 Synovium3 RA	6.9
112386 Ulcer Col-F	1.0	113498 Syn Fluid Cells3 RA	14.3
112738 Match Control Ulcer Col-F	1.9	117106 Normal Cartilage Rep20	2.5
112381 Ulcer Col-M	15.0	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	0.2	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	24.1	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	2.0	117107 Normal Cartilage Rep22	3.4
112383 Ulcer Col-M	6.5	113667 Bone4 Normal	7.7
112736 Match Control Ulcer Col-M	17.4	113668 Synovium4 Normal	8.0
112423 Psoriasis-F	3.9	113669 Syn Fluid Cells4 Normal	8.4

**Table SD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag5929, Run 247683837	Tissue Name	Rel. Exp.(%) Ag5929, Run 247683837
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	1.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	1.2
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0

CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	1.0	Astrocytes rest	19.6
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	2.1
Secondary CD8 lymphocyte act	1.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	3.3
LAK cells rest	2.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.1
LAK cells IL-2	10.1	Liver cirrhosis	3.5
LAK cells IL-2+IL-12	0.0	NCI-H292 none	53.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	44.8
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	100.0
LAK cells PMA/ionomycin	22.2	NCI-H292 IL-13	90.1
NK Cells IL-2 rest	20.9	NCI-H292 IFN gamma	35.8
Two Way MLR 3 day	15.9	HPAEC none	0.0
Two Way MLR 5 day	2.3	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	13.1
PBMC rest	0.6	Lung fibroblast TNF alpha + IL-1 beta	16.6
PBMC PWM	3.8	Lung fibroblast IL-4	4.0
PBMC PHA-L	1.8	Lung fibroblast IL-9	6.4
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	3.8
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	1.2
B lymphocytes CD40L and IL-4	6.5	Dermal fibroblast CCD1070 TNF alpha	1.3
EOL-1 dbcAMP	3.2	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	1.5	Dermal fibroblast IFN gamma	2.4
Dendritic cells none	25.0	Dermal fibroblast IL-4	1.4
Dendritic cells LPS	26.4	Dermal Fibroblasts rest	3.7
Dendritic cells anti-CD40	40.1	Neutrophils TNFa+LPS	0.0
Monocytes rest	5.2	Neutrophils rest	0.0
Monocytes LPS	100.0	Colon	1.3
Macrophages rest	24.5	Lung	2.6
Macrophages LPS	10.5	Thymus	3.5
HUVEC none	0.0	Kidney	9.4
HUVEC starved	0.0		

Table SE. Panel 5D

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag5887, Run 258657708		Ag5887, Run 258657708
97457_Patient-02go_adipose	60.7	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	11.2	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	10.9	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.9
97481_Patient-08sk_skeletal muscle	7.1	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	9.7	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.9	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.8
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	3.6	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	2.7
97492_Patient-10ut_uterus	9.5	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.6	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	100.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	18.8	94735_Donor 3 AD - C_adipose	0.8
97497_Patient-11ut_uterus	24.0	77138_Liver_HepG2untreated	0.9
97498_Patient-11pl_placenta	1.8	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	69.7	81735_Small Intestine	7.6
97501_Patient-12sk_skeletal muscle	24.7	72409_Kidney_Proximal Convoluted Tubule	3.0
97502_Patient-12ut_uterus	5.3	82685_Small intestine_Duodenum	1.4
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	1.0	72410_Kidney_HRCE	14.7
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	6.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.8	73139_Uterus_Uterine smooth muscle cells	0.0

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5929 This gene is widely expressed at low levels in many samples on this panel, with highest expression in normal lung (CT=28).

5 **General\_screening\_panel\_v1.5 Summary:** Ag4480 Highest expression of this gene is seen in a brain cancer cell line (CT=30). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and

melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Ag5887 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 4.1D Summary:** Ag5929 Highest expression of this gene is seen in LPS activated monocytes and IL-9 treated NCI-H292 cells (CTs=31.3). In contrast, expression is undetectable in resting monocytes (CTs=36-37). Lower but substantial levels of expression are found in resting and activated dendritic cells and macrophages. Based on the expression pattern of this transcript, this gene product may be involved in monocyte activation and differentiation. Therefore, antibodies against the protein encoded by this gene may reduce or inhibit inflammation due to monocyte activation or differentiation and be important in the treatment of diseases such as asthma and arthritis. In addition, this transcript is expressed in a cluster of samples derived from NCI-H292 cells. Treatment of these cells does not seem to significantly alter expression of this transcript in this mucoid epidermoid cell line. Thus, the protein could be used to identify certain lung tumors similar to NCI-H292. The encoded protein may also contribute to the normal function of the goblet cells within the lung. Therefore, designing therapeutics to this protein may be important for the treatment of emphysema and asthma as well as other lung diseases in which goblet cells or the mucus they produce have pathological consequences

**Panel 5D Summary:** Ag5887 Highest expression of this gene is seen in adipose (CT=31.3), with low but significant levels of expression detected in a cluster of samples

derived from adipose and skeletal muscle. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

#### **T. CG115540-01: Novel Membrane Protein containing Collagen triple helix repeat**

Expression of gene CG115540-01 was assessed using the primer-probe set Ag4483,  
described in Table TA.

**Table TA. Probe Name Ag4483**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -aatcgatggagagaaggtctct-3'	22	1071	244
Probe	TET-5' -cctttcatttctcttggtgatgccagt-3' -TAMRA	26	1096	245
Reverse	5' -ctgggtctcctttctgtcctt-3'	21	1148	246

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4483 Expression of the CG115540-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**General\_screening\_panel\_v1.4 Summary:** Ag4483 Expression of the CG115540-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**Panel 4.1D Summary:** Ag4483 Expression of the CG115540-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**general oncology screening panel\_v\_2.4 Summary:** Ag4483 Expression of the CG115540-01 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

#### **U. CG118689-01: Uroplakin 1b splice variant**

Expression of gene CG118689-01 was assessed using the primer-probe sets Ag4485 and Ag4484, described in Tables UA and UB. Results of the RTQ-PCR runs are shown in Tables UC, UD, UE and UF.

**Table UA. Probe Name Ag4485**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -tggaaccagctctctcctaata-3'	22	813	247
Probe	TET-5' -tttgtgccccacactaacgtgtgtgt-3' -TAMRA	26	842	248
Reverse	5' -taccatctgacttggcaatgta-3'	22	870	249

**Table UB. Probe Name Ag4484**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -atcacaatcagttttgggttc-3'	21	698	250

Probe	TET-5'-ccatgttctactggagcagaattgaa-3'-TAMRA	26	728	251
Reverse	5'-gaaggaggtatggtggcaac-3'	20	769	252

**Table UC. General screening panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4484, Run 224311245	Rel. Exp.(%) Ag4485, Run 217081575	Rel. Exp.(%) Ag4485, Run 222456670	Rel. Exp.(%) Ag4485, Run 224311257
Adipose	8.8	3.4	3.7	2.7
Melanoma* Hs688(A).T	3.1	0.1	0.1	0.1
Melanoma* Hs688(B).T	2.5	0.1	0.5	0.1
Melanoma* M14	1.8	0.1	0.2	0.0
Melanoma* LOXIMVI	0.9	0.0	0.0	0.0
Melanoma* SK-MEL-5	2.0	0.0	0.0	0.0
Squamous cell carcinoma SCC-4	51.8	21.2	30.8	15.5
Testis Pool	1.5	0.1	0.2	0.2
Prostate ca.* (bone met) PC-3	1.6	0.1	0.1	0.2
Prostate Pool	0.7	0.1	0.1	0.1
Placenta	1.0	1.6	1.7	1.0
Uterus Pool	0.4	0.0	0.1	0.0
Ovarian ca. OVCAR-3	74.7	100.0	100.0	100.0
Ovarian ca. SK-OV-3	12.9	1.1	0.9	1.2
Ovarian ca. OVCAR-4	6.3	4.8	7.6	5.6
Ovarian ca. OVCAR-5	3.1	0.4	0.4	0.1
Ovarian ca. IGROV-1	33.9	14.8	17.1	12.0
Ovarian ca. OVCAR-8	1.5	0.2	0.2	0.1
Ovary	2.6	1.7	2.1	1.5
Breast ca. MCF-7	1.0	0.2	0.1	0.0
Breast ca. MDA-MB-231	1.4	0.1	0.1	0.1
Breast ca. BT 549	2.6	0.0	0.0	0.0
Breast ca. T47D	4.0	0.4	0.5	0.2
Breast ca. MDA-N	0.0	0.0	0.0	0.1
Breast Pool	2.6	0.5	0.4	0.4
Trachea	42.6	20.9	29.7	22.7
Lung	1.3	0.0	0.2	0.2
Fetal Lung	3.0	0.3	0.5	0.3
Lung ca. NCI-N417	0.0	0.0	0.0	0.0
Lung ca. LX-1	0.7	0.0	0.0	0.0
Lung ca. NCI-H146	1.2	0.0	0.1	0.0
Lung ca. SHP-77	0.0	0.1	0.1	0.1
Lung ca. A549	100.0	39.8	52.5	34.9



Lung ca. NCI-H526	0.0	0.0	0.1	0.0
Lung ca. NCI-H23	2.8	0.7	0.8	0.4
Lung ca. NCI-H460	0.6	0.1	0.1	0.1
Lung ca. HOP-62	0.5	0.1	0.1	0.0
Lung ca. NCI-H522	4.0	0.2	0.4	0.1
Liver	0.0	0.0	0.0	0.0
Fetal Liver	1.4	0.4	0.1	0.2
Liver ca. HepG2	2.8	0.3	0.3	0.2
Kidney Pool	16.3	0.9	1.5	0.6
Fetal Kidney	3.6	1.9	1.7	1.2
Renal ca. 786-0	1.2	1.7	1.5	0.8
Renal ca. A498	1.3	0.1	0.1	0.2
Renal ca. ACHN	0.0	0.0	0.2	0.2
Renal ca. UO-31	6.5	3.2	3.2	2.1
Renal ca. TK-10	1.4	0.3	0.1	0.1
Bladder	24.3	0.7	1.1	0.6
Gastric ca. (liver met.) NCI-N87	23.3	7.7	8.2	4.0
Gastric ca. KATO III	0.0	0.1	0.0	0.0
Colon ca. SW-948	1.1	0.0	0.0	0.0
Colon ca. SW480	1.5	0.2	0.3	0.1
Colon ca.* (SW480 met) SW620	0.0	0.1	0.1	0.0
Colon ca. HT29	0.0	0.1	0.1	0.1
Colon ca. HCT-116	0.7	0.2	0.2	0.1
Colon ca. CaCo-2	8.0	1.3	0.7	0.6
Colon cancer tissue	0.0	0.1	0.1	0.0
Colon ca. SW1116	0.8	0.1	0.1	0.0
Colon ca. Colo-205	0.0	0.0	0.0	0.0
Colon ca. SW-48	0.0	0.0	0.1	0.0
Colon Pool	2.1	0.3	0.2	0.2
Small Intestine Pool	6.2	0.4	0.4	0.3
Stomach Pool	3.0	0.7	0.7	0.8
Bone Marrow Pool	2.4	0.1	0.2	0.0
Fetal Heart	6.5	0.7	0.6	0.7
Heart Pool	3.3	0.1	0.2	0.1
Lymph Node Pool	4.6	0.5	0.3	0.6
Fetal Skeletal Muscle	1.5	0.1	0.0	0.1
Skeletal Muscle Pool	0.7	0.1	0.0	0.1
Spleen Pool	0.5	0.1	0.1	0.1
Thymus Pool	8.1	1.1	1.4	1.3

CNS cancer (glio/astro) U87-MG	1.4	0.2	0.2	0.1
CNS cancer (glio/astro) U-118-MG	0.7	0.6	0.1	0.2
CNS cancer (neuro;met) SK-N-AS	0.6	0.2	0.2	0.0
CNS cancer (astro) SF-539	1.5	0.2	0.1	0.1
CNS cancer (astro) SNB-75	2.9	0.3	0.1	0.1
CNS cancer (glio) SNB-19	18.6	15.5	14.7	12.6
CNS cancer (glio) SF-295	2.8	0.3	0.6	0.4
Brain (Amygdala) Pool	5.0	0.2	0.4	0.1
Brain (cerebellum)	0.9	0.1	0.1	0.1
Brain (fetal)	9.0	0.5	0.3	0.4
Brain (Hippocampus) Pool	2.1	0.4	0.2	0.1
Cerebral Cortex Pool	5.8	0.4	0.4	0.2
Brain (Substantia nigra) Pool	2.2	0.2	0.4	0.1
Brain (Thalamus) Pool	5.5	0.3	0.3	0.2
Brain (whole)	1.3	0.1	0.1	0.1
Spinal Cord Pool	2.5	0.6	0.7	0.6
Adrenal Gland	0.7	0.0	0.0	0.1
Pituitary gland Pool	0.5	0.0	0.1	0.0
Salivary Gland	0.0	0.1	0.0	0.1
Thyroid (female)	0.0	0.0	0.0	0.0
Pancreatic ca. CAPAN2	0.0	0.1	0.0	0.1
Pancreas Pool	20.3	2.2	2.5	2.5

Table UD. Oncology\_cell\_line\_screening\_panel\_v3.1

Tissue Name	Rel. Exp.(%) Ag4484, Run 220424640	Rel. Exp.(%) Ag4485, Run 220424639	Tissue Name	Rel. Exp.(%) Ag4484, Run 220424640	Rel. Exp.(%) Ag4485, Run 220424639
Daoy Medulloblastoma/Cerebellum	0.0	0.0	Ca Ski_Cervical epidermoid carcinoma (metastasis)	0.0	0.1
TE671 Medulloblastom/Cerebellum	0.0	0.0	ES-2_Ovarian clear cell carcinoma	0.0	0.0
D283 Med Medulloblastoma/Cerebellum	0.0	0.0	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	0.0	0.0
PFSK-1 Primitive Neuroectodermal/Cerebellum	0.0	0.2	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	0.0	0.0
XF-498_CNS	0.0	0.0	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	0.0	0.0

SNB-78_CNS/glioma	0.0	0.0	Raji_Burkitt's lymphoma	0.5	0.1
SF-268_CNS/glioblastoma	0.0	0.0	Daudi_Burkitt's lymphoma	0.0	0.2
T98G_Glioblastoma	0.0	0.0	U266_B-cell plasmacytoma/myeloma	0.0	0.1
SK-N-SH_Neuroblastoma (metastasis)	0.0	0.0	CA46_Burkitt's lymphoma	0.0	0.0
SF-295_CNS/glioblastoma	0.0	0.0	RL_non-Hodgkin's B-cell lymphoma	0.0	0.0
Cerebellum	1.5	0.2	JM1_pre-B-cell lymphoma/leukemia	0.0	0.0
Cerebellum	0.0	0.1	Jurkat_T cell leukemia	0.0	0.0
NCI-H292_Mucoepidermoid lung ca.	0.0	0.8	TF-1_Erythroleukemia	0.0	0.2
DMS-114_Small cell lung cancer	0.0	0.0	HUT 78_T-cell lymphoma	0.5	0.1
DMS-79_Small cell lung cancer/neuroendocrine	0.0	0.1	U937_Histiocytic lymphoma	0.0	0.1
NCI-H146_Small cell lung cancer/neuroendocrine	0.0	0.1	KU-812_Myelogenous leukemia	0.0	0.0
NCI-H526_Small cell lung cancer/neuroendocrine	0.0	0.1	769-P_Clear cell renal ca.	8.0	14.5
NCI-N417_Small cell lung cancer/neuroendocrine	0.0	0.1	Caki-2_Clear cell renal ca.	2.3	5.0
NCI-H82_Small cell lung cancer/neuroendocrine	0.0	0.1	SW 839_Clear cell renal ca.	0.0	0.1
NCI-H157_Squamous cell lung cancer (metastasis)	0.0	0.0	G401_Wilms' tumor	0.0	0.0
NCI-H1155_Large cell lung cancer/neuroendocrine	0.0	0.0	Hs766T_Pancreatic ca. (LN metastasis)	0.0	0.1
NCI-H1299_Large cell lung cancer/neuroendocrine	0.9	0.0	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	0.0	0.0
NCI-H727_Lung carcinoid	0.0	0.1	SU86.86_Pancreatic carcinoma (liver metastasis)	0.0	0.1
NCI-UMC-11_Lung carcinoid	0.0	0.0	BxPC-3_Pancreatic adenocarcinoma	0.0	0.0
LX-1_Small cell lung cancer	0.0	0.0	HPAC_Pancreatic adenocarcinoma	0.0	0.1
Colo-205_Colon cancer	0.0	0.0	MIA PaCa-2_Pancreatic ca.	0.5	0.5
KM12_Colon cancer	0.0	0.0	CFPAC-1_Pancreatic	6.3	3.3

			ductal adenocarcinoma		
KM20L2_Colon cancer	0.0	0.0	PANC-1_Pancreatic epithelioid ductal ca.	0.0	1.0
NCI-H716_Colon cancer	0.0	0.0	T24_Bladder ca. (transitional cell)	0.0	0.0
SW-48_Colon adenocarcinoma	0.0	0.0	5637_Bladder ca.	0.9	1.2
SW1116_Colon adenocarcinoma	0.0	0.1	HT-1197_Bladder ca.	37.9	100.0
LS 174T_Colon adenocarcinoma	0.0	0.1	UM-UC-3_Bladder ca. (transitional cell)	0.0	0.0
SW-948_Colon adenocarcinoma	0.0	0.0	A204_Rhabdomyosarcoma	0.0	0.0
SW-480_Colon adenocarcinoma	0.0	0.0	HT-1080_Fibrosarcoma	0.0	0.0
NCI-SNU-5_Gastric ca.	0.0	0.0	MG-63_Osteosarcoma (bone)	0.0	0.2
KATO III_Stomach	0.0	0.1	SK-LMS-1_Leiomyosarcoma (vulva)	0.0	0.1
NCI-SNU-16_Gastric ca.	0.0	0.0	SJRH30_Rhabdomyosarcoma (met to bone marrow)	0.0	0.0
NCI-SNU-1_Gastric ca.	4.4	7.7	A431_Epidermoid ca.	7.5	1.4
RF-1_Gastric adenocarcinoma	0.5	0.2	WM266-4_Melanoma	0.0	0.0
RF-48_Gastric adenocarcinoma	0.0	0.0	DU 145_Prostate	0.0	0.0
MKN-45_Gastric ca.	100.0	93.3	MDA-MB-468_Breast adenocarcinoma	0.0	0.3
NCI-N87_Gastric ca.	2.5	5.2	SSC-4_Tongue	6.5	15.6
OVCAR-5_Ovarian ca.	0.0	0.0	SSC-9_Tongue	0.0	0.1
RL95-2_Uterine carcinoma	0.0	0.1	SSC-15_Tongue	0.0	0.0
HelaS3_Cervical adenocarcinoma	0.0	0.0	CAL 27_Squamous cell ca. of tongue	0.0	0.0

**Table UE. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4484, Run 219309629	Tissue Name	Rel. Exp.(%) Ag4484, Run 219309629
Secondary Th1 act	8.0	HUVEC IL-1beta	0.0
Secondary Th2 act	53.6	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	16.3

Secondary Th1 rest	23.2	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	23.8
Secondary Tr1 rest	0.0	Lung Microvascular EC none	11.2
Primary Th1 act	24.7	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	29.7	Microvascular Dermal EC none	0.0
Primary Tr1 act	14.3	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	7.8	Bronchial epithelium TNFalpha + IL1beta	36.9
Primary Th2 rest	12.1	Small airway epithelium none	25.7
Primary Tr1 rest	30.4	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	19.8	Coronary artery SMC rest	11.6
CD45RO CD4 lymphocyte act	30.6	Coronary artery SMC TNFalpha + IL-1beta	15.1
CD8 lymphocyte act	14.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	10.8
Secondary CD8 lymphocyte act	8.7	KU-812 (Basophil) rest	14.6
CD4 lymphocyte none	26.1	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 _anti-CD95 CH11	10.6	CCD1106 (Keratinocytes) none	23.5
LAK cells rest	17.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	12.9
LAK cells IL-2	14.9	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	11.7	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	17.3	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	27.5
LAK cells PMA/ionomycin	26.8	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	29.3	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	64.2	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	10.9
Two Way MLR 7 day	0.0	Lung fibroblast none	18.4
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	9.0
PBMC PWM	12.8	Lung fibroblast IL-4	0.0
PBMC PHA-L	23.2	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	15.5	Dermal fibroblast CCD1070 rest	40.6
B lymphocytes CD40L and IL-4	39.0	Dermal fibroblast CCD1070 TNF alpha	32.8
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	14.4

EOL-1 dbcAMP PMA/ionomycin	11.3	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	19.5
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	8.7	Neutrophils rest	15.4
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	62.4
HUVEC none	0.0	Kidney	79.0
HUVEC starved	0.0		

**Table UF. general oncology screening panel\_v\_2.4**

Tissue Name	Rel. Exp.(%) Ag4484, Run 259807103	Rel. Exp.(%) Ag4485, Run 260280486	Tissue Name	Rel. Exp.(%) Ag4484, Run 259807103	Rel. Exp.(%) Ag4485, Run 260280486
Colon cancer 1	0.7	0.0	Bladder cancer NAT 2	0.0	0.0
Colon cancer NAT 1	0.7	0.1	Bladder cancer NAT 3	0.0	0.0
Colon cancer 2	0.0	0.1	Bladder cancer NAT 4	0.0	0.0
Colon cancer NAT 2	0.0	0.2	Prostate adenocarcinoma 1	1.9	0.6
Colon cancer 3	0.0	0.1	Prostate adenocarcinoma 2	0.0	0.0
Colon cancer NAT 3	0.0	0.2	Prostate adenocarcinoma 3	0.8	0.0
Colon malignant cancer 4	0.0	0.0	Prostate adenocarcinoma 4	0.0	0.1
Colon normal adjacent tissue 4	0.0	0.0	Prostate cancer NAT 5	0.0	0.0
Lung cancer 1	0.0	0.0	Prostate adenocarcinoma 6	0.0	0.0
Lung NAT 1	0.0	0.0	Prostate adenocarcinoma 7	0.0	0.0
Lung cancer 2	9.7	7.4	Prostate adenocarcinoma 8	0.0	0.0
Lung NAT 2	0.0	0.1	Prostate adenocarcinoma 9	0.0	0.2
Squamous cell carcinoma 3	100.0	100.0	Prostate cancer NAT 10	0.0	0.0
Lung NAT 3	0.0	12.9	Kidney cancer 1	0.0	0.2
metastatic melanoma 1	2.3	0.4	Kidney NAT 1	0.0	0.3
Melanoma 2	0.0	0.0	Kidney cancer 2	42.6	34.9
Melanoma 3	0.0	0.0	Kidney NAT 2	1.4	0.3
metastatic melanoma 4	0.0	0.4	Kidney cancer 3	8.9	2.5
metastatic melanoma 5	1.3	0.5	Kidney NAT 3	0.0	0.3
Bladder cancer 1	0.0	0.0	Kidney cancer 4	0.0	0.3
Bladder cancer NAT 1	0.0	0.0	Kidney NAT 4	0.8	0.8
Bladder cancer 2	2.9	3.7			

**General\_screening\_panel\_v1.4 Summary:** Ag4484/Ag4485 Results of four experiments with two different probes and primer sets are in very good agreement with highest expression of the CG118689-01 gene in two cancer cell lines derived from lung and ovarian cancers (CTs=27.9-32). In addition, moderate to low levels of expression of this gene is also seen in number of cancer cell lines derived from ovarian, renal, gastric, squamous cell carcinoma, and brain cancers. Therefore, expression of this gene may be used as diagnostic marker for detection of these cancers and therapeutic modulation of this gene may be beneficial in the treatment of these cancers.

The CG118689-01 gene encodes a splice variant of uroplakin 1B (UPK1B). UPK1B is a structural protein and a terminal differentiation component of the asymmetric unit membrane on the apical surface of the mammalian bladder. UPK1B is a member of the tetraspan family of proteins, many of which have de-regulated patterns of expression in cancer. UPK1B mRNA has been shown to be down-regulated in transitional-cell-bladder-carcinoma and some of the bladder cancer cell lines (Finch et al., 1999, Int J Cancer 80(4):533-8; PMID: 9935153). Therefore, therapeutic modulation of UPK1B protein encoded by this gene may be useful in the treatment of bladder cancer.

Moderate to low levels of expression of this gene is also seen in adipose and pancreas. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Results from three experiments (Ag4484: runs 217218123 and 218333106; Ag4485: run 218981791) with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Oncology\_cell\_line\_screening\_panel\_v3.1 Summary:** Ag4484/Ag4485 Results of two experiments with two different probes and primer sets are in good agreement with highest expression of the CG118689-01 gene in a gastric and bladder cancer cell lines (CTs=28.4-31.7). In addition, moderate to low levels of expression of this gene is also seen in renal and tongue cancer cell lines. Therefore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

**Panel 4.1D Summary:** Ag4485 Highest expression of the CG118689-01 gene is seen in TNFalpha + IL-1beta treated small airway epithelium (CT=33.5). Therefore, expression of this gene may be used to distinguish this sample from other samples in this panel. Furthermore, therapeutic modulation of the expression or activity of the protein encoded by this gene through the application of antibodies or small molecule therapeutics may be useful in the treatment of asthma, COPD, and emphysema.

In addition, low expression of this gene is also seen in kidney samples. Therefore, therapeutic modulation of this gene may be beneficial in the treatment of inflammatory and autoimmune diseases that affect kidney, including lupus erythematosus and glomerulonephritis.

5 Ag4485 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**general oncology screening panel\_v\_2.4 Summary:** Ag4484/Ag4485 Results of two experiments with two different probe and primer sets are in excellent agreements, with highest expression of the CG118689-01 gene in squamous cell carcinoma (CT=27.5-33.7).

10 Moderate to low levels of expression of this gene is also seen kidney and bladder cancer. Expression of this gene is higher in cancer as compared to the corresponding control samples. Therefore, expression of this gene may be used as a diagnostic marker to detect presence of kidney and prostate cancer. Furthermore, therapeutic modulation of the protein encoded by this gene may be useful in the treatment of these cancers. Please see panel 1.4  
15 for further discussion on the utility of this gene.

#### V. CG118689-02: UROPLAKIN IB

Expression of full length physical clone CG118689-02 was assessed using the primer-probe set Ag6811, described in Table VA. Results of the RTQ-PCR runs are shown in Tables VB and VC.

20

**Table VA. Probe Name Ag6811**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - tctgatgtttatagtatatgcctttgaag - 3'	29	283	253
Probe	TET-5' - tggcatccttgatcacagcagcaac - 3' - TAMRA	25	312	254
Reverse	5' - cctctctagcataaagtctcggtt - 3'	25	337	255

**Table VB. General\_screening\_panel\_v1.6**

Tissue Name	Rel. Exp.(%) Ag6811, Run 278018584	Tissue Name	Rel. Exp.(%) Ag6811, Run 278018584
Adipose	2.2	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.2
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	3.4
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.1



Squamous cell carcinoma SCC-4	26.1	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.5
Placenta	1.3	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.1
Ovarian ca. OVCAR-3	100.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	3.8	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.2	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	5.9	Stomach Pool	0.1
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.1
Ovary	1.9	Fetal Heart	0.1
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.1	Thymus Pool	0.4
Trachea	13.9	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	7.6
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	47.6	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.2
Lung ca. NCI-H23	0.4	Brain (fetal)	0.2
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.1	Spinal Cord Pool	0.7
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	1.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.7	Salivary Gland	0.0
Renal ca. A498	0.1	Thyroid (female)	0.0

Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	1.3	Pancreas Pool	0.4

**Table VC. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag6811, Run 278021646	Tissue Name	Rel. Exp.(%) Ag6811, Run 278021646
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	38.2
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	2.5
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	8.4
LAK cells rest	6.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	15.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	5.8
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0

Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	2.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	52.5
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6811 Expression of the CG118689-02 gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**General\_screening\_panel\_v1.6 Summary:** Ag6811 Highest expression of the CG118689-02 gene is detected in ovarian cancer OVCAR-3 cell line (CT=27.3).

In addition, moderate to low levels of expression of this gene is also seen in number of cancer cell lines derived from ovarian, lung, renal, gastric, squamous cell carcinoma, and brain cancers. Therefore, expression of this gene may be used as diagnostic marker for detection of these cancers and therapeutic modulation of this gene may be beneficial in the treatment of these cancers.

The CG118689-01 gene encodes a splice variant of uroplakin 1B (UPK1B). UPK1B is a structural protein and a terminal differentiation component of the asymmetric unit membrane on the apical surface of the mammalian bladder. UPK1B is a member of the tetraspan family of proteins, many of which have de-regulated patterns of expression in cancer. UPK1B mRNA has been shown to be down-regulated in transitional-cell-bladder-carcinoma and some of the bladder cancer cell lines (Finch et al., 1999, Int J Cancer

80(4):533-8; PMID: 9935153). Therefore, therapeutic modulation of UPK1B protein encoded by this gene may be useful in the treatment of bladder cancer.

Moderate to low levels of expression of this gene is also seen in adipose and spinal cord samples. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes and diseases that affect the spinal cord.

**Panel 4.1D Summary:** Ag6811 Highest expression of the CG118689-02 gene is detected in TNFalpha + IL-1beta treated small airway epithelium (CT=33.5). Therefore, expression of this gene may be used to distinguish this sample from other samples in this panel. Furthermore, therapeutic modulation of the expression or activity of the protein encoded by this gene through the application of antibodies or small molecule therapeutics may be useful in the treatment of asthma, COPD, and emphysema.

In addition, low expression of this gene is also seen in kidney samples. Therefore, therapeutic modulation of this gene may be beneficial in the treatment of inflammatory and autoimmune diseases that affect kidney, including lupus erythematosus and glomerulonephritis.

#### W. CG120748-01: LMBR1 LONG FORM

Expression of gene CG120748-01 was assessed using the primer-probe set Ag4507, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB, WC and WD.

**Table WA. Probe Name Ag4507**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-attcatggtttgtggaatcttg-3'	22	328	256
Probe	TET-5'-atttgattgatgccctttgcctttt-3'-TAMRA	26	375	257
Reverse	5'-aaagccttctgattccagaaag-3'	22	402	258

**Table WB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4507, Run 224704541	Tissue Name	Rel. Exp.(%) Ag4507, Run 224704541
AD 1 Hippo	12.1	Control (Path) 3 Temporal Ctx	7.9
AD 2 Hippo	28.5	Control (Path) 4 Temporal Ctx	29.3
AD 3 Hippo	7.4	AD 1 Occipital Ctx	11.9
AD 4 Hippo	6.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	84.1	AD 3 Occipital Ctx	6.3

AD 6 Hippo	22.4	AD 4 Occipital Ctx	16.5
Control 2 Hippo	1.6	AD 5 Occipital Ctx	20.3
Control 4 Hippo	10.2	AD 6 Occipital Ctx	42.0
Control (Path) 3 Hippo	8.7	Control 1 Occipital Ctx	4.8
AD 1 Temporal Ctx	0.0	Control 2 Occipital Ctx	66.9
AD 2 Temporal Ctx	25.7	Control 3 Occipital Ctx	13.5
AD 3 Temporal Ctx	6.3	Control 4 Occipital Ctx	6.5
AD 4 Temporal Ctx	13.4	Control (Path) 1 Occipital Ctx	75.3
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	11.6
AD 5 Sup Temporal Ctx	39.8	Control (Path) 3 Occipital Ctx	0.7
AD 6 Inf Temporal Ctx	39.2	Control (Path) 4 Occipital Ctx	18.6
AD 6 Sup Temporal Ctx	49.3	Control 1 Parietal Ctx	7.5
Control 1 Temporal Ctx	7.7	Control 2 Parietal Ctx	34.2
Control 2 Temporal Ctx	42.0	Control 3 Parietal Ctx	12.8
Control 3 Temporal Ctx	2.9	Control (Path) 1 Parietal Ctx	77.9
Control 4 Temporal Ctx	8.7	Control (Path) 2 Parietal Ctx	25.0
Control (Path) 1 Temporal Ctx	52.1	Control (Path) 3 Parietal Ctx	6.9
Control (Path) 2 Temporal Ctx	32.3	Control (Path) 4 Parietal Ctx	55.1

**Table WC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4507, Run 222695223	Tissue Name	Rel. Exp.(%) Ag4507, Run 222695223
Adipose	8.1	Renal ca. TK-10	36.1
Melanoma* Hs688(A).T	15.6	Bladder	17.6
Melanoma* Hs688(B).T	11.8	Gastric ca. (liver met.) NCI-N87	31.6
Melanoma* M14	36.6	Gastric ca. KATO III	79.0
Melanoma* LOXIMVI	22.1	Colon ca. SW-948	13.2
Melanoma* SK-MEL-5	67.8	Colon ca. SW480	100.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	49.3
Testis Pool	17.4	Colon ca. HT29	22.2
Prostate ca.* (bone met) PC-3	51.1	Colon ca. HCT-116	50.0
Prostate Pool	6.7	Colon ca. CaCo-2	33.4
Placenta	4.4	Colon cancer tissue	31.0
Uterus Pool	3.9	Colon ca. SW1116	9.0
Ovarian ca. OVCAR-3	36.6	Colon ca. Colo-205	6.9
Ovarian ca. SK-OV-3	23.7	Colon ca. SW-48	11.5
Ovarian ca. OVCAR-4	12.3	Colon Pool	16.7
Ovarian ca. OVCAR-5	46.3	Small Intestine Pool	13.8

Ovarian ca. IGROV-1	22.5	Stomach Pool	8.9
Ovarian ca. OVCAR-8	20.2	Bone Marrow Pool	5.2
Ovary	12.2	Fetal Heart	8.8
Breast ca. MCF-7	30.8	Heart Pool	6.1
Breast ca. MDA-MB-231	33.0	Lymph Node Pool	17.8
Breast ca. BT 549	58.6	Fetal Skeletal Muscle	5.9
Breast ca. T47D	87.1	Skeletal Muscle Pool	7.7
Breast ca. MDA-N	18.3	Spleen Pool	10.4
Breast Pool	17.7	Thymus Pool	15.1
Trachea	10.5	CNS cancer (glio/astro) U87-MG	46.0
Lung	4.4	CNS cancer (glio/astro) U-118-MG	32.3
Fetal Lung	29.3	CNS cancer (neuro;met) SK-N-AS	22.7
Lung ca. NCI-N417	15.6	CNS cancer (astro) SF-539	11.6
Lung ca. LX-1	30.8	CNS cancer (astro) SNB-75	40.3
Lung ca. NCI-H146	15.0	CNS cancer (glio) SNB-19	22.7
Lung ca. SHP-77	22.1	CNS cancer (glio) SF-295	74.7
Lung ca. A549	36.3	Brain (Amygdala) Pool	17.6
Lung ca. NCI-H526	11.7	Brain (cerebellum)	11.1
Lung ca. NCI-H23	76.8	Brain (fetal)	24.0
Lung ca. NCI-H460	15.5	Brain (Hippocampus) Pool	18.9
Lung ca. HOP-62	18.7	Cerebral Cortex Pool	30.8
Lung ca. NCI-H522	35.1	Brain (Substantia nigra) Pool	22.2
Liver	2.5	Brain (Thalamus) Pool	26.1
Fetal Liver	11.8	Brain (whole)	23.0
Liver ca. HepG2	10.7	Spinal Cord Pool	13.2
Kidney Pool	20.6	Adrenal Gland	42.0
Fetal Kidney	12.9	Pituitary gland Pool	6.0
Renal ca. 786-0	33.7	Salivary Gland	4.5
Renal ca. A498	10.4	Thyroid (female)	7.1
Renal ca. ACHN	39.2	Pancreatic ca. CAPAN2	52.9
Renal ca. UO-31	27.7	Pancreas Pool	17.2

**Table WD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4507, Run 197487946	Tissue Name	Rel. Exp.(%) Ag4507, Run 197487946
Secondary Th1 act	40.3	HUVEC IL-1beta	28.1
Secondary Th2 act	53.6	HUVEC IFN gamma	29.1
Secondary Tr1 act	34.9	HUVEC TNF alpha + IFN gamma	14.1

Secondary Th1 rest	10.6	HUVEC TNF alpha + IL4	15.2
Secondary Th2 rest	18.0	HUVEC IL-11	17.8
Secondary Tr1 rest	8.5	Lung Microvascular EC none	100.0
Primary Th1 act	17.9	Lung Microvascular EC TNFalpha + IL-1beta	38.7
Primary Th2 act	33.9	Microvascular Dermal EC none	42.3
Primary Tr1 act	29.3	Microvascular Dermal EC TNFalpha + IL-1beta	22.4
Primary Th1 rest	7.9	Bronchial epithelium TNFalpha + IL1beta	25.7
Primary Th2 rest	5.4	Small airway epithelium none	13.3
Primary Tr1 rest	18.7	Small airway epithelium TNFalpha + IL-1beta	27.2
CD45RA CD4 lymphocyte act	25.3	Coronary artery SMC rest	14.7
CD45RO CD4 lymphocyte act	36.6	Coronary artery SMC TNFalpha + IL-1beta	17.1
CD8 lymphocyte act	27.2	Astrocytes rest	13.5
Secondary CD8 lymphocyte rest	37.9	Astrocytes TNFalpha + IL-1beta	16.3
Secondary CD8 lymphocyte act	14.0	KU-812 (Basophil) rest	16.0
CD4 lymphocyte none	12.9	KU-812 (Basophil) PMA/ionomycin	25.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	16.2	CCD1106 (Keratinocytes) none	31.2
LAK cells rest	16.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	15.5
LAK cells IL-2	26.6	Liver cirrhosis	7.0
LAK cells IL-2+IL-12	14.1	NCI-H292 none	33.4
LAK cells IL-2+IFN gamma	17.4	NCI-H292 IL-4	57.4
LAK cells IL-2+ IL-18	21.2	NCI-H292 IL-9	55.5
LAK cells PMA/ionomycin	40.3	NCI-H292 IL-13	55.5
NK Cells IL-2 rest	30.8	NCI-H292 IFN gamma	37.9
Two Way MLR 3 day	29.9	HPAEC none	22.5
Two Way MLR 5 day	28.1	HPAEC TNF alpha + IL-1 beta	34.2
Two Way MLR 7 day	15.2	Lung fibroblast none	13.7
PBMC rest	7.9	Lung fibroblast TNF alpha + IL-1 beta	11.9
PBMC PWM	19.1	Lung fibroblast IL-4	14.8
PBMC PHA-L	17.4	Lung fibroblast IL-9	29.1
Ramos (B cell) none	54.0	Lung fibroblast IL-13	14.9
Ramos (B cell) ionomycin	57.0	Lung fibroblast IFN gamma	16.2
B lymphocytes PWM	25.5	Dermal fibroblast CCD1070 rest	25.7
B lymphocytes CD40L and IL-4	22.8	Dermal fibroblast CCD1070 TNF alpha	46.0
EOL-1 dbcAMP	39.5	Dermal fibroblast CCD1070 IL-1 beta	13.1

EOL-1 dbcAMP PMA/ionomycin	34.2	Dermal fibroblast IFN gamma	10.4
Dendritic cells none	17.9	Dermal fibroblast IL-4	22.7
Dendritic cells LPS	11.3	Dermal Fibroblasts rest	14.7
Dendritic cells anti-CD40	14.5	Neutrophils TNFa+LPS	6.3
Monocytes rest	15.5	Neutrophils rest	16.5
Monocytes LPS	17.4	Colon	3.4
Macrophages rest	17.3	Lung	18.8
Macrophages LPS	8.0	Thymus	34.4
HUVEC none	22.8	Kidney	27.0
HUVEC starved	28.1		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4507 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

5        **General\_screening\_panel\_v1.4 Summary:** Ag4507 Highest expression of this gene is seen in a colon cancer cell line (CT=26). This gene is widely expressed in this panel, with high to moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in  
10 the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that  
15 disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at high to moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful  
20 in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag4507 This gene is expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease, with highest expression in untreated lung microvascular endothelial cells (CT=28.4). In  
25 addition, expression is seen in members of the T-cell, B-cell, endothelial cell,



macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

#### X. CG121519-01: Novel LDL Receptor Domain Containing Protein

Expression of gene CG121519-01 was assessed using the primer-probe set Ag4512, described in Table XA. Results of the RTQ-PCR runs are shown in Tables XB, XC and XD.

**Table XA. Probe Name Ag4512**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -aagccagactgctctgataggt-3'	22	184	259
Probe	TET-5' -acaagcacaacaggaagctgcaattt-3' -TAMRA	26	232	260
Reverse	5' -ccagtttcctgaacttgtttca-3'	22	258	261

**Table XB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4512, Run 224711017	Tissue Name	Rel. Exp.(%) Ag4512, Run 224711017
AD 1 Hippo	31.0	Control (Path) 3 Temporal Ctx	19.1
AD 2 Hippo	94.0	Control (Path) 4 Temporal Ctx	18.8
AD 3 Hippo	2.6	AD 1 Occipital Ctx	24.8
AD 4 Hippo	34.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	1.5
AD 6 Hippo	68.3	AD 4 Occipital Ctx	48.0
Control 2 Hippo	57.0	AD 5 Occipital Ctx	24.5
Control 4 Hippo	29.1	AD 6 Occipital Ctx	12.6
Control (Path) 3 Hippo	14.4	Control 1 Occipital Ctx	10.3
AD 1 Temporal Ctx	27.0	Control 2 Occipital Ctx	100.0

AD 2 Temporal Ctx	49.7	Control 3 Occipital Ctx	33.7
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	7.5
AD 4 Temporal Ctx	46.7	Control (Path) 1 Occipital Ctx	84.7
AD 5 Inf Temporal Ctx	76.3	Control (Path) 2 Occipital Ctx	3.5
AD 5 Sup Temporal Ctx	54.7	Control (Path) 3 Occipital Ctx	1.6
AD 6 Inf Temporal Ctx	42.6	Control (Path) 4 Occipital Ctx	22.5
AD 6 Sup Temporal Ctx	43.5	Control 1 Parietal Ctx	18.8
Control 1 Temporal Ctx	11.3	Control 2 Parietal Ctx	41.8
Control 2 Temporal Ctx	17.0	Control 3 Parietal Ctx	23.7
Control 3 Temporal Ctx	4.2	Control (Path) 1 Parietal Ctx	46.7
Control 3 Temporal Ctx	12.9	Control (Path) 2 Parietal Ctx	21.5
Control (Path) 1 Temporal Ctx	36.1	Control (Path) 3 Parietal Ctx	3.0
Control (Path) 2 Temporal Ctx	29.1	Control (Path) 4 Parietal Ctx	35.8

**Table XC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4512, Run 222696029	Tissue Name	Rel. Exp.(%) Ag4512, Run 222696029
Adipose	5.9	Renal ca. TK-10	25.2
Melanoma* Hs688(A).T	0.1	Bladder	2.8
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.3	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.8	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca. * (SW480 met) SW620	0.0
Testis Pool	9.2	Colon ca. HT29	1.1
Prostate ca.* (bone met) PC-3	0.1	Colon ca. HCT-116	0.0
Prostate Pool	3.8	Colon ca. CaCo-2	28.9
Placenta	0.6	Colon cancer tissue	12.1
Uterus Pool	1.6	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.1
Ovarian ca. OVCAR-4	0.0	Colon Pool	4.8
Ovarian ca. OVCAR-5	0.2	Small Intestine Pool	11.5
Ovarian ca. IGROV-1	0.2	Stomach Pool	8.5
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	2.8
Ovary	6.4	Fetal Heart	2.7
Breast ca. MCF-7	100.0	Heart Pool	3.6
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	7.3

Breast ca. BT 549	0.6	Fetal Skeletal Muscle	2.0
Breast ca. T47D	0.2	Skeletal Muscle Pool	1.0
Breast ca. MDA-N	0.0	Spleen Pool	1.9
Breast Pool	3.2	Thymus Pool	15.4
Trachea	2.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	3.5	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	19.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	9.7
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	3.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	1.4
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	9.5	Brain (fetal)	4.9
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	2.8
Lung ca. HOP-62	0.3	Cerebral Cortex Pool	4.1
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	2.0
Liver	0.0	Brain (Thalamus) Pool	4.5
Fetal Liver	1.4	Brain (whole)	3.1
Liver ca. HepG2	54.0	Spinal Cord Pool	3.2
Kidney Pool	15.9	Adrenal Gland	3.2
Fetal Kidney	20.9	Pituitary gland Pool	4.7
Renal ca. 786-0	0.1	Salivary Gland	0.4
Renal ca. A498	0.0	Thyroid (female)	1.5
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	4.6	Pancreas Pool	3.0

**Table XD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4512, Run 198263643	Tissue Name	Rel. Exp.(%) Ag4512, Run 198263643
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0

Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.5	Liver cirrhosis	2.1
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.7	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	5.8
PBMC rest	0.9	Lung fibroblast TNF alpha + IL-1 beta	3.0
PBMC PWM	0.0	Lung fibroblast IL-4	1.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	2.8
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.7
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	1.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.6
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	2.6
Dendritic cells none	0.8	Dermal fibroblast IL-4	2.9
Dendritic cells LPS	0.7	Dermal Fibroblasts rest	1.2
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.5
Monocytes rest	0.0	Neutrophils rest	1.2

Monocytes LPS	0.0	Colon	100.0
Macrophages rest	0.7	Lung	9.6
Macrophages LPS	0.0	Thymus	14.9
HUVEC none	0.0	Kidney	20.9
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4512 This panel confirms the expression of the CG121519-01 gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

**General\_screening\_panel\_v1.4 Summary:** Ag4512 Highest expression of the CG121519-01 gene is detected in breast cancer MCF-7 cell line (CT=28). In addition, moderate levels of expression of this gene is also detected in number of cancer cell lines derived from lung, renal, liver, colon and brain cancer. Therefore, expression of this gene may be used as diagnostic marker to detect presence of these cancers and also therapeutic modulation of this gene may be useful in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=34.2) when compared to adult liver (CT=40). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal liver suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

In addition, this gene is expressed at moderate to low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders

such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag4512 Highest expression of the CG121519-01 gene is detected in colon (CT=31.1). Low levels of expression of this gene is also seen in lung, kidney and thymus. Expression of this gene seems to be restricted to the normal tissue. Therefore, expression of this gene may be used to distinguish normal tissues represented by colon, lung, kidney and thymus from other samples used in this panel. In addition, therapeutic modulation of this gene may be useful in the treatment of inflammatory and autoimmune disease affecting these tissues including IBD, Crohn's disease, ulcerative colitis, lupus erythematosus, glomerulonephritis, chronic obstructive pulmonary disease, asthma, allergy and emphysema.

#### Y. CG122176-01: Fn domain containing membrane protein

Expression of gene CG122176-01 was assessed using the primer-probe set Ag4524, described in Table YA. Results of the RTQ-PCR runs are shown in Tables YB, YC and YD.

**Table YA. Probe Name Ag4524**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -tcgtggtcctgttcattgtg-3'	19	467	262
Probe	TET-5' -attgccctcttctgccgccagtat-3' -TAMRA	24	496	263
Reverse	5' -ggttcattgtccttgatgatgt-3'	22	521	264

**Table YB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4524, Run 224711018	Tissue Name	Rel. Exp.(%) Ag4524, Run 224711018
AD 1 Hippo	3.8	Control (Path) 3 Temporal Ctx	7.0
AD 2 Hippo	16.4	Control (Path) 4 Temporal Ctx	21.6
AD 3 Hippo	3.8	AD 1 Occipital Ctx	14.5
AD 4 Hippo	4.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	97.3	AD 3 Occipital Ctx	5.4
AD 6 Hippo	14.9	AD 4 Occipital Ctx	16.7
Control 2 Hippo	31.4	AD 5 Occipital Ctx	14.9
Control 4 Hippo	8.6	AD 6 Occipital Ctx	63.3
Control (Path) 3 Hippo	14.1	Control 1 Occipital Ctx	7.1
AD 1 Temporal Ctx	7.2	Control 2 Occipital Ctx	100.0

AD 2 Temporal Ctx	22.4	Control 3 Occipital Ctx	23.3
AD 3 Temporal Ctx	2.5	Control 4 Occipital Ctx	5.0
AD 4 Temporal Ctx	11.1	Control (Path) 1 Occipital Ctx	93.3
AD 5 Inf Temporal Ctx	48.0	Control (Path) 2 Occipital Ctx	16.2
AD 5 Sup Temporal Ctx	23.8	Control (Path) 3 Occipital Ctx	4.8
AD 6 Inf Temporal Ctx	8.1	Control (Path) 4 Occipital Ctx	17.7
AD 6 Sup Temporal Ctx	12.4	Control 1 Parietal Ctx	8.6
Control 1 Temporal Ctx	6.4	Control 2 Parietal Ctx	21.8
Control 2 Temporal Ctx	38.7	Control 3 Parietal Ctx	23.3
Control 3 Temporal Ctx	18.9	Control (Path) 1 Parietal Ctx	65.1
Control 4 Temporal Ctx	7.1	Control (Path) 2 Parietal Ctx	26.2
Control (Path) 1 Temporal Ctx	49.0	Control (Path) 3 Parietal Ctx	4.4
Control (Path) 2 Temporal Ctx	35.1	Control (Path) 4 Parietal Ctx	44.1

**Table YC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4524, Run 222714443	Tissue Name	Rel. Exp.(%) Ag4524, Run 222714443
Adipose	1.6	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.5	Bladder	0.8
Melanoma* Hs688(B).T	1.2	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.2
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.4
Squamous cell carcinoma SCC-4	0.4	Colon ca.* (SW480 met) SW620	0.2
Testis Pool	1.6	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.3	Colon ca. HCT-116	0.0
Prostate Pool	1.6	Colon ca. CaCo-2	0.7
Placenta	0.1	Colon cancer tissue	0.2
Uterus Pool	0.6	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	1.5	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.4	Colon Pool	1.9
Ovarian ca. OVCAR-5	0.2	Small Intestine Pool	2.5
Ovarian ca. IGROV-1	0.3	Stomach Pool	1.9
Ovarian ca. OVCAR-8	0.3	Bone Marrow Pool	0.3
Ovary	3.0	Fetal Heart	4.2
Breast ca. MCF-7	0.2	Heart Pool	8.6
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	2.1

Breast ca. BT 549	3.2	Fetal Skeletal Muscle	19.8
Breast ca. T47D	0.3	Skeletal Muscle Pool	100.0
Breast ca. MDA-N	0.0	Spleen Pool	0.6
Breast Pool	1.7	Thymus Pool	2.1
Trachea	1.1	CNS cancer (glio/astro) U87-MG	0.1
Lung	0.8	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	3.2	CNS cancer (neuro;met) SK-N-AS	1.8
Lung ca. NCI-N417	0.8	CNS cancer (astro) SF-539	0.6
Lung ca. LX-1	0.1	CNS cancer (astro) SNB-75	2.5
Lung ca. NCI-H146	0.2	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	0.2
Lung ca. A549	0.0	Brain (Amygdala) Pool	7.0
Lung ca. NCI-H526	1.0	Brain (cerebellum)	84.1
Lung ca. NCI-H23	0.7	Brain (fetal)	8.4
Lung ca. NCI-H460	0.1	Brain (Hippocampus) Pool	5.4
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	11.2
Lung ca. NCI-H522	0.3	Brain (Substantia nigra) Pool	7.8
Liver	4.3	Brain (Thalamus) Pool	12.2
Fetal Liver	5.1	Brain (whole)	15.2
Liver ca. HepG2	0.1	Spinal Cord Pool	4.6
Kidney Pool	2.3	Adrenal Gland	10.7
Fetal Kidney	10.4	Pituitary gland Pool	1.4
Renal ca. 786-0	0.0	Salivary Gland	10.1
Renal ca. A498	0.0	Thyroid (female)	0.8
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.2
Renal ca. UO-31	0.0	Pancreas Pool	2.1

**Table YD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4524, Run 198263644	Tissue Name	Rel. Exp.(%) Ag4524, Run 198263644
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	1.6
Secondary Tr1 act	1.9	HUVEC TNF alpha + IFN gamma	1.8
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.9
Secondary Th2 rest	0.0	HUVEC IL-11	1.4
Secondary Tr1 rest	0.0	Lung Microvascular EC none	7.5
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	3.7



Primary Th2 act	0.0	Microvascular Dermal EC none	26.6
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	8.5
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	1.9
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	7.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	5.6
CD8 lymphocyte act	0.0	Astrocytes rest	7.6
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	15.2
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	1.6	Liver cirrhosis	10.4
LAK cells IL-2+IL-12	1.8	NCI-H292 none	3.7
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	14.8
LAK cells IL-2+ IL-18	3.6	NCI-H292 IL-9	15.7
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	5.3
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	5.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	2.7
Two Way MLR 7 day	0.0	Lung fibroblast none	3.5
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	3.8
PBMC PHA-L	0.0	Lung fibroblast IL-9	7.4
Ramos (B cell) none	0.0	Lung fibroblast IL-13	4.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	7.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	7.3
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	4.2
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	3.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	100.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	48.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	29.5
Dendritic cells anti-CD40	5.6	Neutrophils TNFa+LPS	1.5
Monocytes rest	0.0	Neutrophils rest	0.9

Monocytes LPS	1.3	Colon	3.5
Macrophages rest	0.0	Lung	25.7
Macrophages LPS	0.0	Thymus	15.7
HUVEC none	0.0	Kidney	36.9
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4524 This panel confirms the expression of this gene at moderate to high levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag4524 Expression of the CG122176-01 gene is highest in skeletal muscle (CT = 26.1). In general, expression of this gene appears to be higher in normal tissues when compared to cancer cell lines. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of cancer.

In addition, this gene is expressed at high to moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4.1D Summary:** Ag4524 Expression of the CG122176-01 gene is highest in dermal fibroblasts treated with interferon gamma (CT = 32.5) or IL-4 (CT = 33.5).

Therefore, expression of this gene could be used to distinguish dermal fibroblasts from the other samples on this panel. Furthermore, expression of this gene in treated dermal fibroblasts suggests that this gene product may be involved in skin disorders, including psoriasis. In addition, low levels of expression are seen in normal lung and kidney tissues.

**Z. CG122691-01: Fn3/TSPN/Collagen domain containing protein**

Expression of gene CG122691-01 was assessed using the primer-probe set Ag4531, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB, ZC and ZD.

**Table ZA. Probe Name Ag4531**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - actcggaacaggaggtgatact - 3'	22	207	265
Probe	TET- 5' - accaccaagaccctaaggccacagt - 3' - TAMRA	26	230	266
Reverse	5' - ctcggaagatctgcaaggtgtag - 3'	22	280	267

5

**Table ZB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag4531, Run 224711128	Tissue Name	Rel. Exp.(%) Ag4531, Run 224711128
AD 1 Hippo	22.7	Control (Path) 3 Temporal Ctx	17.2
AD 2 Hippo	40.3	Control (Path) 4 Temporal Ctx	24.0
AD 3 Hippo	16.7	AD 1 Occipital Ctx	10.0
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	92.7	AD 3 Occipital Ctx	22.4
AD 6 Hippo	59.9	AD 4 Occipital Ctx	19.6
Control 2 Hippo	55.1	AD 5 Occipital Ctx	0.0
Control 4 Hippo	36.3	AD 6 Occipital Ctx	0.0
Control (Path) 3 Hippo	4.4	Control 1 Occipital Ctx	17.3
AD 1 Temporal Ctx	13.2	Control 2 Occipital Ctx	62.0
AD 2 Temporal Ctx	17.1	Control 3 Occipital Ctx	27.7
AD 3 Temporal Ctx	11.7	Control 4 Occipital Ctx	46.3
AD 4 Temporal Ctx	31.9	Control (Path) 1 Occipital Ctx	61.6
AD 5 Inf Temporal Ctx	90.1	Control (Path) 2 Occipital Ctx	16.8
AD 5 Sup Temporal Ctx	24.0	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	64.6	Control (Path) 4 Occipital Ctx	6.8
AD 6 Sup Temporal Ctx	23.7	Control 1 Parietal Ctx	17.8
Control 1 Temporal Ctx	17.7	Control 2 Parietal Ctx	43.2
Control 2 Temporal Ctx	43.2	Control 3 Parietal Ctx	22.1
Control 3 Temporal Ctx	14.2	Control (Path) 1 Parietal Ctx	91.4
Control 4 Temporal Ctx	29.3	Control (Path) 2 Parietal Ctx	55.9
Control (Path) 1 Temporal Ctx	100.0	Control (Path) 3 Parietal Ctx	22.5
Control (Path) 2 Temporal Ctx	62.9	Control (Path) 4 Parietal Ctx	43.8

**Table ZC. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4531, Run 222735216	Tissue Name	Rel. Exp.(%) Ag4531, Run 222735216
Adipose	1.9	Renal ca. TK-10	8.5
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	2.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	2.0
Testis Pool	2.5	Colon ca. HT29	1.3
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.9	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	1.5
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	1.3
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	2.8
Ovarian ca. IGROV-1	1.6	Stomach Pool	0.0
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	1.4
Breast ca. MCF-7	0.0	Heart Pool	1.6
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	2.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	1.7
Breast ca. T47D	4.8	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	1.6	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	1.4	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	3.6	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	1.5
Lung ca. A549	2.0	Brain (Amygdala) Pool	23.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	20.4

Lung ca. NCI-H23	0.0	Brain (fetal)	100.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	19.3
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	17.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	50.0
Liver	0.0	Brain (Thalamus) Pool	37.4
Fetal Liver	0.0	Brain (whole)	13.2
Liver ca. HepG2	0.0	Spinal Cord Pool	18.9
Kidney Pool	1.8	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	4.1	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**Table ZD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4531, Run 198383969	Tissue Name	Rel. Exp.(%) Ag4531, Run 198383969
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.0

CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	4.8
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	1.9
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	1.7
Macrophages LPS	0.0	Thymus	2.7
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4531 This panel does not show differential expression of this gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

- 5 **General\_screening\_panel\_v1.4 Summary:** Ag4531 Expression of this gene is seen exclusive to the brain in this panel, with highest expression in the fetal brain (CT=32.7). Thus, expression of this gene could be used to differentiate between brain tissue and non-

neuronal tissue. Furthermore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag4531 Expression of this gene is exclusive to the kidney in this panel (CT=32.9). Thus, expression of this gene could be used to differentiate the kidney derived sample from other samples on this panel and as a marker of kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

#### 10 AA. CG122863-01 and CG122863-02: Novel Membrane Protein

Expression of gene CG122863-01 and full length physical clone CG122863-02 was assessed using the primer-probe set Ag4542, described in Table AAA. Results of the RTQ-PCR runs are shown in Tables AAB and AAC. Please note that CG122863-02 represents a full-length physical clone of the CG122863-01 gene, validating the prediction of the gene sequence.

**Table AAA. Probe Name Ag4542**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - caaagtacaggggtgcaagtaa - 3'	22	141	268
Probe	TET- 5' - catcataaataacttcacatctgctgga - 3' - TAMRA	29	163	269
Reverse	5' - gaaagacaggaagaggacgatt - 3'	22	199	270

**Table AAB. General\_screening\_panel\_v1.4**

Tissue Name	Rel. Exp.(%) Ag4542, Run 222809443	Tissue Name	Rel. Exp.(%) Ag4542, Run 222809443
Adipose	10.3	Renal ca. TK-10	14.8
Melanoma* Hs688(A).T	5.7	Bladder	26.4
Melanoma* Hs688(B).T	13.0	Gastric ca. (liver met.) NCI-N87	9.5
Melanoma* M14	7.0	Gastric ca. KATO III	5.1
Melanoma* LOXIMVI	3.8	Colon ca. SW-948	3.2
Melanoma* SK-MEL-5	1.3	Colon ca. SW480	11.5
Squamous cell carcinoma SCC-4	1.5	Colon ca.* (SW480 met) SW620	3.2
Testis Pool	52.9	Colon ca. HT29	0.8
Prostate ca.* (bone met) PC-3	13.9	Colon ca. HCT-116	2.6

Prostate Pool	1.1	Colon ca. CaCo-2	8.2
Placenta	0.8	Colon cancer tissue	3.7
Uterus Pool	0.7	Colon ca. SW1116	3.1
Ovarian ca. OVCAR-3	12.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	69.7	Colon ca. SW-48	1.0
Ovarian ca. OVCAR-4	6.1	Colon Pool	2.8
Ovarian ca. OVCAR-5	11.3	Small Intestine Pool	4.1
Ovarian ca. IGROV-1	3.0	Stomach Pool	1.7
Ovarian ca. OVCAR-8	6.0	Bone Marrow Pool	2.0
Ovary	4.7	Fetal Heart	2.0
Breast ca. MCF-7	0.7	Heart Pool	0.4
Breast ca. MDA-MB-231	18.4	Lymph Node Pool	4.0
Breast ca. BT 549	100.0	Fetal Skeletal Muscle	4.2
Breast ca. T47D	19.8	Skeletal Muscle Pool	1.3
Breast ca. MDA-N	9.9	Spleen Pool	7.4
Breast Pool	3.3	Thymus Pool	8.6
Trachea	5.3	CNS cancer (glio/astro) U87-MG	1.5
Lung	1.2	CNS cancer (glio/astro) U-118-MG	36.6
Fetal Lung	21.0	CNS cancer (neuro;met) SK-N-AS	0.6
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	2.0
Lung ca. LX-1	1.5	CNS cancer (astro) SNB-75	7.6
Lung ca. NCI-H146	1.8	CNS cancer (glio) SNB-19	2.6
Lung ca. SHP-77	6.7	CNS cancer (glio) SF-295	3.0
Lung ca. A549	3.0	Brain (Amygdala) Pool	2.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.0
Lung ca. NCI-H23	4.2	Brain (fetal)	3.0
Lung ca. NCI-H460	1.3	Brain (Hippocampus) Pool	2.1
Lung ca. HOP-62	1.9	Cerebral Cortex Pool	3.0
Lung ca. NCI-H522	33.0	Brain (Substantia nigra) Pool	3.2
Liver	0.3	Brain (Thalamus) Pool	3.2
Fetal Liver	11.4	Brain (whole)	2.8
Liver ca. HepG2	8.5	Spinal Cord Pool	2.8
Kidney Pool	4.4	Adrenal Gland	34.4
Fetal Kidney	4.6	Pituitary gland Pool	0.0
Renal ca. 786-0	13.5	Salivary Gland	3.1
Renal ca. A498	15.8	Thyroid (female)	5.2
Renal ca. ACHN	18.6	Pancreatic ca. CAPAN2	4.7
Renal ca. UO-31	26.2	Pancreas Pool	5.1



**Table AAC. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag4542, Run 198395781	Tissue Name	Rel. Exp.(%) Ag4542, Run 198395781
Secondary Th1 act	18.8	HUVEC IL-1beta	17.3
Secondary Th2 act	12.9	HUVEC IFN gamma	5.2
Secondary Tr1 act	8.0	HUVEC TNF alpha + IFN gamma	35.4
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	25.5
Secondary Th2 rest	1.4	HUVEC IL-11	3.7
Secondary Tr1 rest	2.1	Lung Microvascular EC none	27.4
Primary Th1 act	22.5	Lung Microvascular EC TNFalpha + IL-1beta	39.8
Primary Th2 act	19.5	Microvascular Dermal EC none	26.6
Primary Tr1 act	21.2	Microvascular Dermal EC TNFalpha + IL-1beta	28.7
Primary Th1 rest	3.0	Bronchial epithelium TNFalpha + IL1beta	2.5
Primary Th2 rest	0.9	Small airway epithelium none	2.4
Primary Tr1 rest	1.4	Small airway epithelium TNFalpha + IL-1beta	12.1
CD45RA CD4 lymphocyte act	5.6	Coronary artery SMC rest	6.9
CD45RO CD4 lymphocyte act	7.1	Coronary artery SMC TNFalpha + IL-1beta	7.9
CD8 lymphocyte act	5.7	Astrocytes rest	4.2
Secondary CD8 lymphocyte rest	3.2	Astrocytes TNFalpha + IL-1beta	14.2
Secondary CD8 lymphocyte act	5.0	KU-812 (Basophil) rest	1.7
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	12.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.2	CCD1106 (Keratinocytes) none	1.6
LAK cells rest	10.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	3.3
LAK cells IL-2	4.8	Liver cirrhosis	1.1
LAK cells IL-2+IL-12	5.8	NCI-H292 none	8.4
LAK cells IL-2+IFN gamma	2.2	NCI-H292 IL-4	7.9
LAK cells IL-2+ IL-18	2.4	NCI-H292 IL-9	9.4
LAK cells PMA/ionomycin	88.3	NCI-H292 IL-13	8.7
NK Cells IL-2 rest	3.4	NCI-H292 IFN gamma	20.7
Two Way MLR 3 day	4.8	HPAEC none	6.8
Two Way MLR 5 day	5.1	HPAEC TNF alpha + IL-1 beta	100.0
Two Way MLR 7 day	5.3	Lung fibroblast none	4.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	12.0

PBMC PWM	4.9	Lung fibroblast IL-4	2.8
PBMC PHA-L	4.0	Lung fibroblast IL-9	3.7
Ramos (B cell) none	0.0	Lung fibroblast IL-13	3.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	4.8
B lymphocytes PWM	3.3	Dermal fibroblast CCD1070 rest	2.4
B lymphocytes CD40L and IL-4	1.8	Dermal fibroblast CCD1070 TNF alpha	4.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	2.3
EOL-1 dbcAMP PMA/ionomycin	7.7	Dermal fibroblast IFN gamma	0.7
Dendritic cells none	7.7	Dermal fibroblast IL-4	2.0
Dendritic cells LPS	19.5	Dermal Fibroblasts rest	2.1
Dendritic cells anti-CD40	4.7	Neutrophils TNFa+LPS	0.6
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	8.2	Colon	0.0
Macrophages rest	7.7	Lung	2.1
Macrophages LPS	13.7	Thymus	2.5
HUVEC none	4.7	Kidney	5.2
HUVEC starved	9.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4542 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

**General\_screening\_panel\_v1.4 Summary:** Ag4542 Expression of the CG122863-01 gene is highest in a breast cancer cell line (CT = 31). Furthermore,

- 5 CG122863-01 gene expression appears to be upregulated in a subset of breast and renal cell cancer cell lines when compared to the normal tissue controls. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of renal and breast cancer.

- 10 Among tissues with metabolic or endocrine function, this gene is expressed at low levels in adipose, adrenal gland, and fetal liver. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

- 15 Interestingly, this gene is expressed at higher levels in fetal liver and lung when compared to adult tissues. Therefore, expression of this gene may be used to distinguish adult and fetal liver or lung.

**Panel 4.1D Summary:** Ag4542 Expression of the CG122863-01 gene is highest in HPAECs treated with TNF alpha and IL-1 beta (CT = 30.9).

In general, this transcript is expressed at higher levels in endothelial cells. IL-1 beta and TNFalpha treatment increases transcript levels consistently in endothelium samples including HPAEC, HUVEC, microvascular dermal EC and lung microvascular EC. Therefore, therapeutic modulation of the activity of this gene or its protein product could be important in regulating endothelium function including leukocyte extravasation, a major component of inflammation during asthma, IBD, and psoriasis.

Moderate expression of this gene is detected in PMA/ionomycin treated lymphokine-activated killer cells (LAK). These cells are involved in tumor immunology and cell clearance of virally and bacterial infected cells as well as tumors. Therefore, modulation of the function of the protein encoded by this gene through the application of a small molecule drug or antibody may alter the functions of these cells and lead to improvement of symptoms associated with these conditions.

#### AB. CG50880-04: NEUROTTRIMIN

Expression of full length physical clone CG50880-04 was assessed using the primer-probe set Ag93, described in Table ABA. Results of the RTQ-PCR runs are shown in Tables ABB, ABC, ABD and ABE.

**Table ABA. Probe Name Ag93**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atcctcgcgtgggtccttct-3'	19	360	271
Probe	TET-5'-cacccaaacgcagtagcagcatcgagat-3'-TAMRA	27	327	272
Reverse	5'-tcgtcatcacatccacgttctg-3'	23	303	273

**Table ABB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag93, Run 271695183	Tissue Name	Rel. Exp.(%) Ag93, Run 271695183
AD 1 Hippo	17.2	Control (Path) 3 Temporal Ctx	12.7
AD 2 Hippo	42.6	Control (Path) 4 Temporal Ctx	53.6
AD 3 Hippo	12.9	AD 1 Occipital Ctx	38.4
AD 4 Hippo	20.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	16.6
AD 6 Hippo	52.1	AD 4 Occipital Ctx	50.3
Control 2 Hippo	40.3	AD 5 Occipital Ctx	30.4
Control 4 Hippo	25.2	AD 6 Occipital Ctx	75.3
Control (Path) 3 Hippo	8.5	Control 1 Occipital Ctx	12.5

AD 1 Temporal Ctx	30.8	Control 2 Occipital Ctx	81.8
AD 2 Temporal Ctx	39.0	Control 3 Occipital Ctx	44.4
AD 3 Temporal Ctx	12.9	Control 4 Occipital Ctx	19.2
AD 4 Temporal Ctx	44.4	Control (Path) 1 Occipital Ctx	76.8
AD 5 Inf Temporal Ctx	79.6	Control (Path) 2 Occipital Ctx	28.9
AD 5 Sup Temporal Ctx	44.1	Control (Path) 3 Occipital Ctx	4.6
AD 6 Inf Temporal Ctx	46.0	Control (Path) 4 Occipital Ctx	45.7
AD 6 Sup Temporal Ctx	59.0	Control 1 Parietal Ctx	21.5
Control 1 Temporal Ctx	19.8	Control 2 Parietal Ctx	46.3
Control 2 Temporal Ctx	43.8	Control 3 Parietal Ctx	30.6
Control 3 Temporal Ctx	32.8	Control (Path) 1 Parietal Ctx	87.1
Control 4 Temporal Ctx	26.4	Control (Path) 2 Parietal Ctx	48.6
Control (Path) 1 Temporal Ctx	74.2	Control (Path) 3 Parietal Ctx	9.0
Control (Path) 2 Temporal Ctx	63.3	Control (Path) 4 Parietal Ctx	71.7

**Table ABC. Panel 1**

Tissue Name	Rel. Exp.(%) Ag93, Run 87586352	Rel. Exp.(%) Ag93, Run 88706692	Tissue Name	Rel. Exp.(%) Ag93, Run 87586352	Rel. Exp.(%) Ag93, Run 88706692
Endothelial cells	0.2	0.7	Renal ca. 786-0	0.9	1.5
Endothelial cells (treated)	0.0	0.2	Renal ca. A498	0.0	0.2
Pancreas	0.2	0.4	Renal ca. RXF 393	0.2	0.4
Pancreatic ca. CAPAN 2	0.0	0.2	Renal ca. ACHN	0.0	0.2
Adrenal gland	1.0	0.8	Renal ca. UO-31	8.4	12.7
Thyroid	0.2	0.4	Renal ca. TK-10	0.0	0.1
Salivary gland	0.2	0.6	Liver	0.1	0.3
Pituitary gland	0.1	0.3	Liver (fetal)	0.1	0.3
Brain (fetal)	6.8	7.4	Liver ca. (hepatoblast) HepG2	0.0	0.2
Brain (whole)	25.0	23.5	Lung	0.0	0.2
Brain (amygdala)	11.1	6.1	Lung (fetal)	0.9	1.2
Brain (cerebellum)	100.0	100.0	Lung ca. (small cell) LX-1	0.0	0.2
Brain (hippocampus)	23.2	7.1	Lung ca. (small cell) NCI- H69	2.1	1.4
Brain (substantia nigra)	7.9	9.3	Lung ca. (s.cell var.) SHP- 77	0.0	0.3
Brain (thalamus)	10.8	14.0	Lung ca. (large cell) NCI- H460	0.0	2.7
Brain (hypothalamus)	0.1	0.3	Lung ca. (non-sm. cell) A549	0.0	0.2

Spinal cord	4.5	7.3	Lung ca. (non-s.cell) NCI-H23	0.0	0.2
glio/astro U87-MG	0.0	0.2	Lung ca. (non-s.cell) HOP-62	0.8	0.8
glio/astro U-118-MG	0.7	1.0	Lung ca. (non-s.cl) NCI-H522	0.0	0.2
astrocytoma SW1783	0.1	0.3	Lung ca. (squam.) SW 900	0.0	0.2
neuro*; met SK-N-AS	0.0	0.2	Lung ca. (squam.) NCI-H596	3.6	3.6
astrocytoma SF-539	0.2	0.7	Mammary gland	0.6	1.8
astrocytoma SNB-75	1.7	3.5	Breast ca.* (pl.ef) MCF-7	0.0	0.2
glioma SNB-19	1.0	2.4	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.2
glioma U251	0.6	1.1	Breast ca.* (pl. ef) T47D	0.1	0.2
glioma SF-295	0.2	0.8	Breast ca. BT-549	0.0	0.2
Heart	0.9	1.0	Breast ca. MDA-N	0.0	0.2
Skeletal muscle	0.1	0.3	Ovary	0.1	0.4
Bone marrow	0.1	0.3	Ovarian ca. OVCAR-3	0.0	0.2
Thymus	0.3	0.8	Ovarian ca. OVCAR-4	0.0	0.2
Spleen	0.0	0.2	Ovarian ca. OVCAR-5	0.0	0.3
Lymph node	0.2	0.4	Ovarian ca. OVCAR-8	0.0	0.2
Colon (ascending)	0.3	0.8	Ovarian ca. IGROV-1	0.0	0.2
Stomach	0.6	1.5	Ovarian ca. (ascites) SK-OV-3	0.0	0.2
Small intestine	1.4	1.9	Uterus	0.2	0.3
Colon ca. SW480	0.1	0.3	Placenta	0.0	0.2
Colon ca.* SW620 (SW480 met)	0.0	0.2	Prostate	0.1	0.5
Colon ca. HT29	0.0	0.2	Prostate ca.* (bone met) PC-3	0.0	0.2
Colon ca. HCT-116	0.0	0.2	Testis	0.6	1.3
Colon ca. CaCo-2	0.0	0.2	Melanoma Hs688(A).T	1.1	1.0
Colon ca. HCT-15	0.0	0.2	Melanoma* (met) Hs688(B).T	2.8	3.4
Colon ca. HCC-2998	0.0	0.2	Melanoma UACC-62	0.3	0.6
Gastric ca.* (liver met) NCI-N87	0.0	0.1	Melanoma M14	1.7	1.8
Bladder	0.5	1.5	Melanoma LOX IMVI	18.0	23.0
Trachea	0.4	0.9	Melanoma* (met) SK-MEL-5	0.0	0.3
Kidney	0.1	0.3	Melanoma SK-MEL-28	0.2	0.4
Kidney (fetal)	2.0	2.6			

**Table ABD. Panel 1.3D**

Tissue Name	Rel. Exp.(%) Ag93, Run 165517577	Tissue Name	Rel. Exp.(%) Ag93, Run 165517577
Liver adenocarcinoma	1.0	Kidney (fetal)	2.0
Pancreas	0.2	Renal ca. 786-0	2.2
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.2
Adrenal gland	1.2	Renal ca. RXF 393	0.9
Thyroid	0.3	Renal ca. ACHN	0.0
Salivary gland	0.4	Renal ca. UO-31	19.1
Pituitary gland	0.1	Renal ca. TK-10	0.0
Brain (fetal)	28.3	Liver	0.1
Brain (whole)	82.9	Liver (fetal)	0.2
Brain (amygdala)	33.9	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	100.0	Lung	8.4
Brain (hippocampus)	37.6	Lung (fetal)	1.3
Brain (substantia nigra)	31.2	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	35.6	Lung ca. (small cell) NCI-H69	1.9
Cerebral Cortex	42.9	Lung ca. (s.cell var.) SHP-77	0.1
Spinal cord	17.6	Lung ca. (large cell) NCI-H460	11.3
glio/astro U87-MG	0.1	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	13.5	Lung ca. (non-s.cell) NCI-H23	0.1
astrocytoma SW1783	0.5	Lung ca. (non-s.cell) HOP-62	4.3
neuro*; met SK-N-AS	0.1	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	1.1	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.7	Lung ca. (squam.) NCI-H596	12.4
glioma SNB-19	3.3	Mammary gland	2.8
glioma U251	20.3	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.8	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	1.2	Breast ca.* (pl.ef) T47D	0.1
Heart	2.9	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	3.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.3	Ovary	0.1
Bone marrow	0.2	Ovarian ca. OVCAR-3	0.1
Thymus	0.1	Ovarian ca. OVCAR-4	0.0
Spleen	0.2	Ovarian ca. OVCAR-5	0.0
Lymph node	0.5	Ovarian ca. OVCAR-8	0.0
Colorectal	3.4	Ovarian ca. IGROV-1	0.0
Stomach	1.1	Ovarian ca.* (ascites) SK-OV-3	0.0

Small intestine	8.3	Uterus	0.4
Colon ca. SW480	0.5	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.2
Colon ca. HT29	0.1	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.2	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.9
Colon ca. tissue(ODO3866)	6.2	Melanoma* (met) Hs688(B).T	3.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	4.7
Gastric ca.* (liver met) NCI-N87	0.3	Melanoma M14	10.4
Bladder	2.5	Melanoma LOX IMVI	6.4
Trachea	1.4	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.3	Adipose	2.1

**Table ABE. Panel 4D**

Tissue Name	Rel. Exp.(%) Ag93, Run 164183830	Tissue Name	Rel. Exp.(%) Ag93, Run 164183830
Secondary Th1 act	0.0	HUVEC IL-1beta	0.7
Secondary Th2 act	0.0	HUVEC IFN gamma	1.7
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	2.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	3.3
Secondary Th2 rest	0.0	HUVEC IL-11	0.8
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.5
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	3.7
Primary Th2 act	0.0	Microvascular Dermal EC none	0.4
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.2
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	35.4
Primary Th2 rest	0.0	Small airway epithelium none	17.6
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	50.3
CD45RA CD4 lymphocyte act	12.0	Coronary artery SMC rest	59.9
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	37.1
CD8 lymphocyte act	0.0	Astrocytes rest	31.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	44.4
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	68.8

LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	13.0
LAK cells IL-2	0.0	Liver cirrhosis	6.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.4
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	1.1
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.2
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.6
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.6
Two Way MLR 5 day	0.0	HPAEC none	3.2
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	2.0
PBMC rest	0.2	Lung fibroblast none	56.3
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	44.4
PBMC PHA-L	0.0	Lung fibroblast IL-4	100.0
Ramos (B cell) none	0.1	Lung fibroblast IL-9	98.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	76.8
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	87.7
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	65.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	65.1
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	31.4
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	27.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	38.4
Dendritic cells anti-CD40	0.0	IBD Colitis 2	1.0
Monocytes rest	0.1	IBD Crohn's	3.1
Monocytes LPS	0.0	Colon	16.4
Macrophages rest	0.0	Lung	14.7
Macrophages LPS	0.0	Thymus	1.8
HUVEC none	3.5	Kidney	1.9
HUVEC starved	4.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag93 This panel confirms the expression of the CG50880-04 gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly down-regulated in the temporal cortex of Alzheimer's disease patients. Therefore, up-regulation of this gene or its protein product, or treatment with specific agonists for this receptor may be of use in reversing the dementia, memory loss, and neuronal death associated with this disease.

**Panel 1 Summary:** Ag93 Results of two experiments with same probe and primer sets are in excellent agreements, with highest expression of the CG50880-04 gene in cerebellum (CTs=23). In addition, this gene is expressed at high levels in all regions of the



central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebral cortex, and spinal cord. This gene codes for a splice variant of neurotrimin (Ntm), which belongs to IgLON family of neural cell adhesion molecules. Ntm plays a role in the development of thalamocortical and pontocerebellar projections. It mediates homophilic adhesion and promotes the outgrowth of DRG neurons. However, both membrane-bound and soluble Ntm inhibit the outgrowth of sympathetic neurons (Gil et al., 1998, J Neurosci 18(22):9312-25, PMID: 9801370). Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Moderate to low levels of expression of this gene is also seen in number of cancer cell lines derived from melanoma, lung, renal and brain cancers. Therefore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma, lung, renal and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 1.3D Summary:** Ag93 Highest expression of the CG50880-04 gene is detected in cerebellum (CT=26.5). In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebral cortex, and spinal cord. Moderate to low levels of expression of this gene is also seen in number of cancer cell lines derived from melanoma, lung, renal, colon, liver and brain cancers. Therefore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma, lung, renal, colon, liver and brain cancers. Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Please see panel 1 for further discussion on the utility of this gene.

**Panel 2D Summary:** Ag93 Results from one experiment with the CG50880-04 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 4D Summary:** Ag93 Highest expression of the CG50880-04 gene is detected in cytokine treated lung fibroblasts (CTs=27.5). High expression of this gene is seen in cytokine treated and untreated lung fibroblasts and dermal fibroblasts. Therefore, expression of this gene may be used to distinguish fibroblasts from other samples used in this panel. In addition, moderate to low levels of expression of this gene is also detected in endothelial cells including HUVEC, HPAEC, lung microvascular endothelial cells, TNFalpha + IL1beta treated bronchial epithelium cells, small airway epithelium, astrocytes, keratinocytes, cytokine treated NCI-H292 cells, liver cirrhosis and lupus kidney samples and normal tissues represented by lung, colon, thymus and kidney. Therefore, therapeutic modulation of the Ntm protein encoded by this gene may be useful in the treatment of inflammatory and autoimmune diseases that affect colon, kidney, lung, heart and brain including psoriasis, asthma, allergies, chronic obstructive pulmonary disease, emphysema, inflammatory bowel diseases such as Crohn's and ulcerative colitis, lupus erythematosus, glomerulonephritis and liver cirrhosis.

In addition, moderate expression of this gene is also seen activated CD45RA CD4 lymphocyte (CT=30.5), which represent activated naive T cells. In activated memory T cells (CD45RO CD4 lymphocyte) or CD4 Th1 or Th2 cells, resting CD4 cells the expression of this gene is strongly down regulated (CTs=40) suggesting a role for this putative protein in differentiation or activation of naive T cells. Therefore, modulation of the expression and/or activity of this Ntm protein encoded by this gene might be beneficial for the control of autoimmune diseases and T cell mediated diseases such as arthritis, psoriasis, IBD and asthma.

#### **AC. CG51923-01 and CG51923-02: PROTOCADHERIN FAT2**

Expression of gene CG51923-01 and variant CG51923-02 was assessed using the primer-probe sets Ag395, Ag706, Ag888, Ag944 and Ag945, described in Tables ACA, ACB, ACC, ACD and ACE. Results of the RTQ-PCR runs are shown in Tables ACF, ACG, ACH, ACI, ACJ, ACK and ACL.

**Table ACA. Probe Name Ag395**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -caggaagaaataagccaagtcca-3'	23	13104	274
Probe	TET-5' -tccttggcctcccgcctgc-3' -TAMRA	19	13084	275
Reverse	5' -gaggtcatgttctagcttcccatt-3'	24	13049	276

**Table ACB. Probe Name Ag706**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -tatgtggagagcttcgagaaaa-3'	22	164	277
Probe	TET-5' -atctacctcgcgagccacagtg-3' -TAMRA	23	191	278
Reverse	5' -agagatgatccgtacctcact-3'	22	217	279

**Table ACC. Probe Name Ag888**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -catagctgaccgcatctgaa-3'	20	11160	280
Probe	TET-5' -aatgctccatctccttggtgagtga-3' -TAMRA	26	11125	281
Reverse	5' -ggagctagcatccatcatcac-3'	21	11104	282

**Table ACD. Probe Name Ag944**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -agctcaactacagcaccactgt-3'	22	10296	283
Probe	TET-5' -cagcaaagtccctgcagctgatcctg-3' -TAMRA	25	10339	284
Reverse	5' -ctctggagaatctgggtcact-3'	21	10364	285

**Table ACE. Probe Name Ag945**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ccaagtcattcatgtcaga-3'	22	5581	286
Probe	TET-5' -ttccctccagattctcagaacaga-3' -TAMRA	26	5614	287
Reverse	5' -atggataggcccgactattg-3'	20	5652	288

5

**Table ACF. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag888, Run 224758716	Rel. Exp.(%) Ag888, Run 268770555	Tissue Name	Rel. Exp.(%) Ag888, Run 224758716	Rel. Exp.(%) Ag888, Run 268770555
AD 1 Hippo	3.1	1.9	Control (Path) 3 Temporal Ctx	21.2	0.0
AD 2 Hippo	29.7	7.0	Control (Path) 4 Temporal Ctx	62.0	24.7
AD 3 Hippo	0.0	1.7	AD 1 Occipital Ctx	25.0	9.3
AD 4 Hippo	8.9	8.0	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	99.3	92.7	AD 3 Occipital Ctx	1.9	6.1
AD 6 Hippo	20.4	22.4	AD 4 Occipital Ctx	24.5	18.2
Control 2 Hippo	14.3	5.4	AD 5 Occipital Ctx	20.6	27.0
Control 4 Hippo	14.8	7.1	AD 6 Occipital Ctx	26.6	11.5
Control (Path) 3 Hippo	6.8	1.9	Control 1 Occipital Ctx	3.7	4.4
AD 1 Temporal Ctx	11.7	13.5	Control 2 Occipital Ctx	44.4	55.5
AD 2 Temporal Ctx	20.4	27.7	Control 3 Occipital Ctx	40.6	20.2

AD 3 Temporal Ctx	4.8	3.4	Control 4 Occipital Ctx	13.2	6.7
AD 4 Temporal Ctx	34.9	30.1	Control (Path) 1 Occipital Ctx	69.7	56.3
AD 5 Inf Temporal Ctx	100.0	100.0	Control (Path) 2 Occipital Ctx	20.6	15.8
AD 5 SupTemporal Ctx	23.3	36.1	Control (Path) 3 Occipital Ctx	0.0	1.7
AD 6 Inf Temporal Ctx	73.7	39.5	Control (Path) 4 Occipital Ctx	36.6	33.7
AD 6 Sup Temporal Ctx	75.3	54.0	Control 1 Parietal Ctx	10.7	3.7
Control 1 Temporal Ctx	9.9	5.4	Control 2 Parietal Ctx	49.3	33.0
Control 2 Temporal Ctx	16.8	19.2	Control 3 Parietal Ctx	35.6	27.4
Control 3 Temporal Ctx	36.1	11.5	Control (Path) 1 Parietal Ctx	49.3	55.9
Control 4 Temporal Ctx	6.7	6.6	Control (Path) 2 Parietal Ctx	35.6	25.3
Control (Path) 1 Temporal Ctx	81.2	31.9	Control (Path) 3 Parietal Ctx	2.7	4.5
Control (Path) 2 Temporal Ctx	59.5	54.7	Control (Path) 4 Parietal Ctx	44.1	44.8

**Table ACG. Panel 1.1**

Tissue Name	Rel. Exp.(%) Ag395, Run 109668522	Tissue Name	Rel. Exp.(%) Ag395, Run 109668522
Adrenal gland	0.1	Renal ca. UO-31	0.1
Bladder	1.4	Renal ca. RXF 393	0.0
Brain (amygdala)	0.1	Liver	0.5
Brain (cerebellum)	100.0	Liver (fetal)	0.5
Brain (hippocampus)	0.2	Liver ca. (hepatoblast) HepG2	0.0
Brain (substantia nigra)	1.2	Lung	0.1
Brain (thalamus)	0.2	Lung (fetal)	0.2
Cerebral Cortex	1.5	Lung ca. (non-s.cell) HOP-62	1.0
Brain (fetal)	0.9	Lung ca. (large cell) NCI-H460	0.8
Brain (whole)	4.5	Lung ca. (non-s.cell) NCI-H23	0.2
glio/astro U-118-MG	0.1	Lung ca. (non-s.cl) NCI-H522	0.7
astrocytoma SF-539	0.3	Lung ca. (non-sm. cell) A549	0.3
astrocytoma SNB-75	0.3	Lung ca. (s.cell var.) SHP-77	0.2
astrocytoma SW1783	0.1	Lung ca. (small cell) LX-1	1.2
glioma U251	0.1	Lung ca. (small cell) NCI-H69	0.4
glioma SF-295	0.4	Lung ca. (squam.) SW 900	0.1

glioma SNB-19	0.1	Lung ca. (squam.) NCI-H596	0.5
glio/astro U87-MG	0.8	Lymph node	0.3
neuro*; met SK-N-AS	1.2	Spleen	0.1
Mammary gland	1.4	Thymus	1.1
Breast ca. BT-549	0.2	Ovary	0.0
Breast ca. MDA-N	0.7	Ovarian ca. IGROV-1	0.1
Breast ca.* (pl.ef) T47D	0.5	Ovarian ca. OVCAR-3	7.7
Breast ca.* (pl.ef) MCF-7	0.3	Ovarian ca. OVCAR-4	6.4
Breast ca.* (pl.ef) MDA-MB-231	0.1	Ovarian ca. OVCAR-5	1.5
Small intestine	0.6	Ovarian ca. OVCAR-8	0.5
Colorectal	0.2	Ovarian ca.* (ascites) SK-OV-3	0.7
Colon ca. HT29	0.1	Pancreas	0.9
Colon ca. CaCo-2	1.0	Pancreatic ca. CAPAN 2	0.0
Colon ca. HCT-15	0.3	Pituitary gland	0.5
Colon ca. HCT-116	0.3	Placenta	0.6
Colon ca. HCC-2998	1.1	Prostate	2.4
Colon ca. SW480	0.3	Prostate ca.* (bone met) PC-3	0.2
Colon ca.* SW620 (SW480 met)	1.0	Salivary gland	2.4
Stomach	0.3	Trachea	1.9
Gastric ca. (liver met) NCI-N87	0.5	Spinal cord	0.4
Heart	0.4	Testis	2.0
Skeletal muscle (Fetal)	0.5	Thyroid	0.1
Skeletal muscle	0.8	Uterus	0.1
Endothelial cells	0.2	Melanoma M14	0.4
Heart (Fetal)	0.0	Melanoma LOX IMVI	0.1
Kidney	0.7	Melanoma UACC-62	0.1
Kidney (fetal)	0.7	Melanoma SK-MEL-28	1.6
Renal ca. 786-0	0.1	Melanoma* (met) SK-MEL-5	0.1
Renal ca. A498	0.3	Melanoma Hs688(A).T	0.1
Renal ca. ACHN	0.3	Melanoma* (met) Hs688(B).T	0.1
Renal ca. TK-10	0.5		

**Table ACH. Panel 1.2**

Tissue Name	Rel. Exp.(%) Ag706, Run 116351409	Rel. Exp.(%) Ag706, Run 118348190	Rel. Exp.(%) Ag888, Run 118840724	Rel. Exp.(%) Ag888, Run 119442398
Endothelial cells	0.0	0.0	0.0	0.0
Heart (Fetal)	0.0	0.0	0.0	0.0
Pancreas	0.4	0.4	0.2	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	0.0	0.0

Adrenal Gland	0.2	0.1	0.0	0.0
Thyroid	0.5	0.3	0.0	0.0
Salivary gland	4.8	4.6	8.8	2.7
Pituitary gland	0.7	0.7	0.5	0.0
Brain (fetal)	0.8	0.7	0.7	0.0
Brain (whole)	9.0	17.0	22.7	20.2
Brain (amygdala)	0.3	0.3	0.5	0.0
Brain (cerebellum)	100.0	96.6	100.0	100.0
Brain (hippocampus)	0.4	0.4	0.4	0.0
Brain (thalamus)	0.1	0.1	0.2	0.0
Cerebral Cortex	0.5	0.8	2.7	0.1
Spinal cord	0.1	0.2	0.2	0.0
glio/astro U87-MG	0.3	0.4	0.0	0.0
glio/astro U-118-MG	0.1	0.0	0.0	0.0
astrocytoma SW1783	0.0	0.0	0.0	0.0
neuro*; met SK-N-AS	0.0	0.0	0.0	0.0
astrocytoma SF-539	0.0	0.0	0.0	0.0
astrocytoma SNB-75	0.2	0.1	0.2	0.0
glioma SNB-19	0.0	0.1	0.0	0.0
glioma U251	0.2	0.1	0.0	0.0
glioma SF-295	0.0	0.1	0.0	0.0
Heart	0.1	0.1	0.0	0.0
Skeletal Muscle	1.2	0.8	0.1	0.0
Bone marrow	0.0	0.0	0.3	0.0
Thymus	0.3	0.5	0.8	0.0
Spleen	0.0	0.0	0.0	0.0
Lymph node	0.1	0.1	0.2	0.0
Colorectal Tissue	0.1	0.0	0.1	0.0
Stomach	0.2	0.3	0.3	0.0
Small intestine	0.1	0.1	0.0	0.0
Colon ca. SW480	0.0	0.0	0.0	0.0
Colon ca.* SW620 (SW480 met)	0.1	0.0	0.1	0.0
Colon ca. HT29	0.1	0.0	0.0	0.0
Colon ca. HCT-116	0.0	0.0	0.0	0.0
Colon ca. CaCo-2	0.1	0.1	0.0	0.0
Colon ca. Tissue (ODO3866)	0.2	0.1	0.0	0.0
Colon ca. HCC-2998	0.0	0.0	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.2	0.2	0.1	0.0
Bladder	0.5	0.5	1.3	0.0

Trachea	1.6	2.7	3.7	1.2
Kidney	0.9	0.5	0.4	0.0
Kidney (fetal)	1.4	1.2	1.7	0.2
Renal ca. 786-0	0.2	0.3	0.0	0.0
Renal ca. A498	1.6	2.2	0.1	0.0
Renal ca. RXF 393	0.0	0.0	0.0	0.0
Renal ca. ACHN	0.1	0.1	0.0	0.0
Renal ca. UO-31	0.0	0.0	0.0	0.0
Renal ca. TK-10	0.4	0.6	0.0	0.0
Liver	0.1	0.0	0.0	0.0
Liver (fetal)	0.0	0.0	0.0	0.0
Liver ca. (hepatoblast) HepG2	0.0	0.0	0.0	0.0
Lung	0.0	0.1	0.0	0.0
Lung (fetal)	0.0	0.1	0.0	0.0
Lung ca. (small cell) LX-1	0.1	0.1	0.3	0.0
Lung ca. (small cell) NCI-H69	0.2	0.2	1.4	0.0
Lung ca. (s.cell var.) SHP-77	0.0	0.0	0.0	0.0
Lung ca. (large cell) NCI-H460	0.1	0.1	0.1	0.0
Lung ca. (non-sm. cell) A549	0.1	0.1	0.1	0.0
Lung ca. (non-s.cell) NCI-H23	0.0	0.0	0.0	0.0
Lung ca. (non-s.cell) HOP-62	0.1	0.1	0.0	0.0
Lung ca. (non-s.cl) NCI-H522	0.0	0.0	0.0	0.0
Lung ca. (squamous) SW 900	0.0	0.1	0.0	0.0
Lung ca. (squamous) NCI-H596	0.1	0.1	0.7	0.0
Mammary gland	2.0	2.7	5.8	2.9
Breast ca.* (pl.ef) MCF-7	0.0	0.0	0.0	0.0
Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0	0.0	0.0
Breast ca.* (pl. ef) T47D	0.3	0.7	0.2	0.0
Breast ca. BT-549	0.0	0.1	0.1	0.0
Breast ca. MDA-N	0.2	0.3	0.0	0.0
Ovary	0.1	0.2	0.0	0.0
Ovarian ca. OVCAR-3	61.6	63.3	29.3	16.3
Ovarian ca. OVCAR-4	62.0	100.0	35.6	22.2
Ovarian ca. OVCAR-5	0.7	0.6	0.5	0.0
Ovarian ca. OVCAR-8	0.1	0.1	0.1	0.0
Ovarian ca. IGROV-1	0.1	0.1	0.0	0.0
Ovarian ca. (ascites) SK-OV-3	0.7	0.9	0.3	0.0
Uterus	0.0	0.0	0.0	0.0
Placenta	1.5	1.7	1.1	0.2

Prostate	1.8	1.7	3.8	0.6
Prostate ca.* (bone met) PC-3	0.0	0.0	0.0	0.0
Testis	4.4	6.5	20.6	10.5
Melanoma Hs688(A).T	0.0	0.0	0.0	0.0
Melanoma* (met) Hs688(B).T	0.1	0.0	0.0	0.0
Melanoma UACC-62	0.2	0.1	0.0	0.0
Melanoma M14	0.3	0.7	0.1	0.0
Melanoma LOX IMVI	0.0	0.0	0.0	0.0
Melanoma* (met) SK-MEL-5	1.3	1.8	0.2	0.0

**Table ACl. Panel 1.3D**

Tissue Name	Rel. Exp.(%) Ag888, Run 166006522	Tissue Name	Rel. Exp.(%) Ag888, Run 166006522
Liver adenocarcinoma	0.0	Kidney (fetal)	0.1
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.2
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.2
Salivary gland	1.4	Renal ca. UO-31	0.0
Pituitary gland	0.3	Renal ca. TK-10	0.1
Brain (fetal)	0.1	Liver	0.0
Brain (whole)	19.5	Liver (fetal)	0.0
Brain (amygdala)	0.1	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	100.0	Lung	0.0
Brain (hippocampus)	0.2	Lung (fetal)	0.0
Brain (substantia nigra)	0.4	Lung ca. (small cell) LX-1	0.1
Brain (thalamus)	0.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.2	Lung ca. (s.cell var.) SHP-77	0.1
Spinal cord	0.1	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	0.2	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.2	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	1.2
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.1	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.1



Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	5.2
Thymus	1.0	Ovarian ca. OVCAR-4	15.4
Spleen	0.0	Ovarian ca. OVCAR-5	0.1
Lymph node	0.0	Ovarian ca. OVCAR-8	0.1
Colorectal	0.1	Ovarian ca. IGROV-1	0.0
Stomach	0.2	Ovarian ca. * (ascites) SK-OV-3	0.1
Small intestine	0.1	Uterus	0.0
Colon ca. SW480	0.0	Placenta	1.4
Colon ca.* SW620(SW480 met)	0.1	Prostate	0.6
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	3.5
Colon ca. CaCo-2	0.1	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.1	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.1
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	1.0	Melanoma* (met) SK-MEL-5	0.2
Kidney	0.2	Adipose	0.0

**Table ACJ. Panel 2D**

Tissue Name	Rel. Exp.(%) Ag395, Run 144794701	Rel. Exp.(%) Ag888, Run 144791434	Rel. Exp.(%) Ag888, Run 145420466	Tissue Name	Rel. Exp.(%) Ag395, Run 144794701	Rel. Exp.(%) Ag888, Run 144791434	Rel. Exp.(%) Ag888, Run 145420466
Normal Colon	20.2	10.7	5.6	Kidney Margin 8120608	1.1	0.4	1.9
CC Well to Mod Diff (ODO3866)	6.0	0.5	0.5	Kidney Cancer 8120613	0.9	0.0	0.4
CC Margin (ODO3866)	5.8	0.0	0.0	Kidney Margin 8120614	2.0	0.0	0.3
CC Gr.2 rectosigmoid (ODO3868)	1.8	0.7	0.2	Kidney Cancer 9010320	13.1	2.4	0.9
CC Margin (ODO3868)	1.9	0.6	0.7	Kidney Margin 9010321	11.5	2.3	3.1
CC Mod Diff (ODO3920)	2.2	2.0	0.7	Normal Uterus	2.9	0.1	1.1
CC Margin (ODO3920)	5.6	1.1	1.1	Uterus Cancer 064011	21.3	23.2	21.2

CC Gr.2 ascend colon (ODO3921)	1.2	0.3	0.0	Normal Thyroid	0.8	0.7	0.7
CC Margin (ODO3921)	0.9	0.8	0.9	Thyroid Cancer 064010	2.5	3.2	1.5
CC from Partial Hepatectomy (ODO4309) Mets	0.9	0.7	0.2	Thyroid Cancer A302152	3.0	0.7	1.5
Liver Margin (ODO4309)	1.3	0.9	0.0	Thyroid Margin A302153	0.0	0.4	0.5
Colon mets to lung (OD04451-01)	2.2	0.7	0.4	Normal Breast	44.1	9.2	16.6
Lung Margin (OD04451-02)	5.4	0.6	0.2	Breast Cancer (OD04566)	5.3	1.7	3.8
Normal Prostate 6546-1	43.8	29.3	21.0	Breast Cancer (OD04590-01)	10.8	1.2	1.2
Prostate Cancer (OD04410)	17.3	9.3	5.2	Breast Cancer Mets (OD04590-03)	6.4	3.1	3.3
Prostate Margin (OD04410)	15.7	8.9	12.2	Breast Cancer Metastasis (OD04655-05)	1.4	0.2	1.4
Prostate Cancer (OD04720-01)	41.2	37.9	41.2	Breast Cancer 064006	13.1	13.7	5.5
Prostate Margin (OD04720-02)	22.8	37.1	33.2	Breast Cancer 1024	62.0	55.9	23.3
Normal Lung 061010	2.8	4.5	3.0	Breast Cancer 9100266	10.0	22.4	14.4
Lung Met to Muscle (ODO4286)	0.0	1.3	1.3	Breast Margin 9100265	12.9	36.6	28.5
Muscle Margin (ODO4286)	66.0	24.0	16.7	Breast Cancer A209073	25.2	43.8	44.8
Lung Malignant Cancer (OD03126)	3.5	4.4	2.4	Breast Margin A209073	61.1	100.0	20.7
Lung Margin (OD03126)	2.9	1.8	0.2	Normal Liver	5.4	0.0	0.4
Lung Cancer (OD04404)	46.0	100.0	30.4	Liver Cancer 064003	2.6	1.0	0.0
Lung Margin (OD04404)	16.6	5.9	1.7	Liver Cancer 1025	1.0	0.4	0.3
Lung Cancer (OD04565)	100.0	65.5	100.0	Liver Cancer 1026	0.9	0.0	0.0

Lung Margin (OD04565)	3.0	0.8	2.0	Liver Cancer 6004-T	9.7	0.0	0.0
Lung Cancer (OD04237-01)	2.6	0.9	1.2	Liver Tissue 6004-N	3.1	0.6	0.2
Lung Margin (OD04237-02)	0.6	0.9	0.2	Liver Cancer 6005-T	0.0	0.3	0.3
Ocular Mel Met to Liver (ODO4310)	1.0	0.7	0.9	Liver Tissue 6005-N	0.0	0.0	0.0
Liver Margin (ODO4310)	0.0	0.0	0.0	Normal Bladder	9.0	2.5	3.1
Melanoma Mets to Lung (OD04321)	3.5	1.1	0.3	Bladder Cancer 1023	2.4	0.4	0.3
Lung Margin (OD04321)	0.8	1.2	0.5	Bladder Cancer A302173	21.8	33.4	11.9
Normal Kidney	11.3	10.3	3.0	Bladder Cancer (OD04718-01)	46.7	75.3	68.3
Kidney Ca, Nuclear grade 2 (OD04338)	6.3	2.3	2.4	Bladder Normal Adjacent (OD04718-03)	4.1	1.6	0.5
Kidney Margin (OD04338)	3.6	3.4	1.4	Normal Ovary	0.0	0.4	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	23.8	4.5	3.3	Ovarian Cancer 064008	65.1	91.4	50.3
Kidney Margin (OD04339)	15.0	5.8	5.5	Ovarian Cancer (OD04768-07)	33.0	17.9	10.8
Kidney Ca, Clear cell type (OD04340)	3.2	1.0	2.4	Ovary Margin (OD04768-08)	0.0	0.0	0.2
Kidney Margin (OD04340)	11.9	9.8	4.0	Normal Stomach	2.4	2.1	1.6
Kidney Ca, Nuclear grade 3 (OD04348)	1.3	2.0	1.9	Gastric Cancer 9060358	1.5	0.7	0.0
Kidney Margin (OD04348)	12.2	2.7	3.0	Stomach Margin 9060359	1.4	0.4	0.4
Kidney Cancer (OD04622-01)	4.9	2.5	4.8	Gastric Cancer 9060395	2.3	0.4	0.2
Kidney Margin (OD04622-03)	3.1	3.2	4.4	Stomach Margin 9060394	0.8	0.3	0.7
Kidney Cancer	0.5	1.6	0.1	Gastric Cancer	6.6	2.8	0.8

(OD04450-01)				9060397			
Kidney Margin (OD04450-03)	7.4	3.0	0.8	Stomach Margin 9060396	0.0	0.0	0.2
Kidney Cancer 8120607	3.0	2.7	0.4	Gastric Cancer 064005	4.5	1.5	0.3

**Table ACK. Panel 3D**

Tissue Name	Rel. Exp.(%) Ag395, Run 164730889	Tissue Name	Rel. Exp.(%) Ag395, Run 164730889
Daoy- Medulloblastoma	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	4.8
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.5
D283 Med- Medulloblastoma	0.8	Ramos- Stimulated with PMA/ionomycin 6h	0.3
PFSK-1- Primitive Neuroectodermal	0.3	Ramos- Stimulated with PMA/ionomycin 14h	0.1
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megakaryoblast)	0.5
SNB-78- Glioma	0.3	Raji- Burkitt's lymphoma	1.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.8
T98G- Glioblastoma	0.3	U266- B-cell plasmacytoma	0.2
SK-N-SH- Neuroblastoma (metastasis)	0.2	CA46- Burkitt's lymphoma	0.1
SF-295- Glioblastoma	0.2	RL- non-Hodgkin's B-cell lymphoma	0.1
Cerebellum	84.1	JM1- pre-B-cell lymphoma	0.3
Cerebellum	100.0	Jurkat- T cell leukemia	0.2
NCI-H292- Mucoepidermoid lung carcinoma	2.0	TF-1- Erythroleukemia	0.2
DMS-114- Small cell lung cancer	0.2	HUT 78- T-cell lymphoma	0.4
DMS-79- Small cell lung cancer	3.0	U937- Histiocytic lymphoma	0.6
NCI-H146- Small cell lung cancer	1.1	KU-812- Myelogenous leukemia	0.7
NCI-H526- Small cell lung cancer	0.6	769-P- Clear cell renal carcinoma	0.3
NCI-N417- Small cell lung cancer	0.1	Caki-2- Clear cell renal carcinoma	0.5
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.1
NCI-H157- Squamous cell lung cancer (metastasis)	0.2	G401- Wilms' tumor	0.1
NCI-H1155- Large cell lung cancer	1.9	Hs766T- Pancreatic carcinoma (LN metastasis)	0.5
NCI-H1299- Large cell lung cancer	0.7	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.1

NCI-H727- Lung carcinoid	0.1	SU86.86- Pancreatic carcinoma (liver metastasis)	0.2
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	12.1
LX-1- Small cell lung cancer	0.3	HPAC- Pancreatic adenocarcinoma	0.1
Colo-205- Colon cancer	0.5	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.4	CFPAC-1- Pancreatic ductal adenocarcinoma	0.1
KM20L2- Colon cancer	0.1	PANC-1- Pancreatic epithelioid ductal carcinoma	0.2
NCI-H716- Colon cancer	0.2	T24- Bladder carcinoma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.7	5637- Bladder carcinoma	9.0
SW1116- Colon adenocarcinoma	0.3	HT-1197- Bladder carcinoma	0.3
LS 174T- Colon adenocarcinoma	0.5	UM-UC-3- Bladder carcinoma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.2	HT-1080- Fibrosarcoma	0.4
NCI-SNU-5- Gastric carcinoma	0.4	MG-63- Osteosarcoma	0.4
KATO III- Gastric carcinoma	0.5	SK-LMS-1- Leiomyosarcoma (vulva)	0.6
NCI-SNU-16- Gastric carcinoma	0.4	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.3
NCI-SNU-1- Gastric carcinoma	0.1	A431- Epidermoid carcinoma	25.5
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.3
RF-48- Gastric adenocarcinoma	0.3	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.2	MDA-MB-468- Breast adenocarcinoma	1.3
NCI-N87- Gastric carcinoma	0.1	SCC-4- Squamous cell carcinoma of tongue	0.2
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	5.9	SCC-15- Squamous cell carcinoma of tongue	0.1
HelaS3- Cervical adenocarcinoma	0.7	CAL 27- Squamous cell carcinoma of tongue	38.4

**Table ACL. Panel CNS\_1**

Tissue Name	Rel. Exp.(%) Ag888, Run 171791108	Tissue Name	Rel. Exp.(%) Ag888, Run 171791108
BA4 Control	26.1	BA17 PSP	0.0
BA4 Control2	6.4	BA17 PSP2	20.4
BA4 Alzheimer's2	0.0	Sub Nigra Control	11.5

BA4 Parkinson's	50.3	Sub Nigra Control2	11.2
BA4 Parkinson's2	52.1	Sub Nigra Alzheimer's2	23.2
BA4 Huntington's	0.0	Sub Nigra Parkinson's2	43.8
BA4 Huntington's2	80.1	Sub Nigra Huntington's	69.3
BA4 PSP	19.5	Sub Nigra Huntington's2	27.4
BA4 PSP2	11.2	Sub Nigra PSP2	0.0
BA4 Depression	0.0	Sub Nigra Depression	0.0
BA4 Depression2	11.2	Sub Nigra Depression2	0.0
BA7 Control	0.0	Glob Palladus Control	0.0
BA7 Control2	43.2	Glob Palladus Control2	62.9
BA7 Alzheimer's2	13.5	Glob Palladus Alzheimer's	0.0
BA7 Parkinson's	27.0	Glob Palladus Alzheimer's2	11.1
BA7 Parkinson's2	60.3	Glob Palladus Parkinson's	64.6
BA7 Huntington's	44.8	Glob Palladus Parkinson's2	38.7
BA7 Huntington's2	73.2	Glob Palladus PSP	11.3
BA7 PSP	49.7	Glob Palladus PSP2	0.0
BA7 PSP2	38.2	Glob Palladus Depression	36.1
BA7 Depression	21.0	Temp Pole Control	0.0
BA9 Control	0.0	Temp Pole Control2	25.3
BA9 Control2	43.5	Temp Pole Alzheimer's	10.5
BA9 Alzheimer's	0.0	Temp Pole Alzheimer's2	11.2
BA9 Alzheimer's2	36.6	Temp Pole Parkinson's	0.0
BA9 Parkinson's	55.1	Temp Pole Parkinson's2	39.0
BA9 Parkinson's2	23.2	Temp Pole Huntington's	14.3
BA9 Huntington's	37.9	Temp Pole PSP	0.0
BA9 Huntington's2	50.0	Temp Pole PSP2	0.0
BA9 PSP	26.6	Temp Pole Depression2	2.9
BA9 PSP2	12.4	Cing Gyr Control	16.3
BA9 Depression	0.0	Cing Gyr Control2	19.5
BA9 Depression2	34.2	Cing Gyr Alzheimer's	0.0
BA17 Control	100.0	Cing Gyr Alzheimer's2	10.3
BA17 Control2	10.7	Cing Gyr Parkinson's	47.6
BA17 Alzheimer's2	18.2	Cing Gyr Parkinson's2	14.6
BA17 Parkinson's	48.3	Cing Gyr Huntington's	40.9
BA17 Parkinson's2	54.3	Cing Gyr Huntington's2	14.1
BA17 Huntington's	13.8	Cing Gyr PSP	11.3
BA17 Huntington's2	43.5	Cing Gyr PSP2	15.4
BA17 Depression	12.9	Cing Gyr Depression	26.8
BA17 Depression2	34.9	Cing Gyr Depression2	34.6

**CNS\_neurodegeneration\_v1.0 Summary:** Ag888 Results of two experiments with same probe and primer sets are in good agreement. This panel confirms the expression of the CG51923-01 gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's  
5 diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.1 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**Panel 1.1 Summary:** Ag395 Highest expression of the CG51923-01 gene is detected in cerebellum (CT=21). Therefore, expression of this gene may be used to  
10 differentiate cerebellum from other samples used in this panel. In addition, high to moderate levels of expression of this gene is also seen in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebral cortex, and spinal cord. This gene codes for protocadherin Fat 2 protein, a homolog of the *Drosophila* tumor suppressor gene *fat*. Protocadherins are transmembrane glycoproteins  
15 belonging to the cadherin superfamily of molecules, which are involved in many biological processes such as cell adhesion, cytoskeletal organization and morphogenesis. Protocadherins generally exhibit only moderate adhesive activity and are highly expressed in the nervous system. FAT2 occupies an isolated position in the cadherin superfamily, because they contain EGF domains together with the classical cadherin repeats (Nollet et  
20 al., 2000, J Mol Biol 299(3):551-72, PMID: 10835267). Cadherins can act as axon guidance and cell adhesion proteins, specifically during development and in the response to injury (Ranscht B., 2000, Int. J. Dev. Neurosci. 18: 643-651, PMID: 10978842). Therefore, manipulation of levels of this protein may be of use in inducing a compensatory synaptogenic response to neuronal death in Alzheimer's disease, Parkinson's disease,  
25 Huntington's disease, spinocerebellar ataxia, progressive supranuclear palsy, ALS, head trauma, stroke, or any other disease/condition associated with neuronal loss.

Moderate to high levels of expression of this gene is also seen in cluster of cancer cell lines derived from gastric, colon, lung, renal, breast, ovarian, prostate, melanoma and brain cancers. Thus, therapeutic modulation of the expression or function of this gene may  
30 be effective in the treatment of gastric, colon, lung, renal, breast, ovarian, prostate, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity

of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 1.2 Summary:** Ag706/Ag888 Results of four experiments with two different probe and primer sets are in very good agreement. Highest expression of the CG51923-01 gene is detected in cerebellum and a ovarian cancer cell line (CTs=23-25). In addition, high to moderate levels of expression of this gene is also seen in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebral cortex, and spinal cord, in cluster of cancer cell lines derived from gastric, colon, lung, renal, breast, ovarian, prostate, melanoma and brain cancers. Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Please see panel 1.1 for further discussion on utility of this gene.

**Panel 1.3D Summary:** Ag888 Highest expression of the CG51923-01 gene is detected in cerebellum (CT=27). In addition, moderate levels of expression of this gene is also seen in two cancer cell lines derived from ovarian cancer. Please see panel 1.1 for further discussion on utility of this gene.

**Panel 2D Summary:** Ag395/Ag888 Results of three experiments with two different probe and primer sets are in good agreements. Highest expression of the CG51923-01 gene is detected in two lung cancer cell lines and a control breast sample (CTs=29-32). Moderate levels of expression of this gene is also seen in samples derived from ovarian, bladder, breast, uterine, lung, and prostate cancers. Expression of this gene is higher in ovarian, bladder and lung cancers as compared to their corresponding control samples. Therefore, expression of this gene may be used as diagnostic marker for detection of these cancers. Furthermore, therapeutic modulation of the protocadherin encoded by this gene through the use of antibodies or small molecule drug may be beneficial in the treatment of ovarian, bladder, breast, uterine, lung, and prostate cancers.

**Panel 3D Summary:** Ag395 Highest expression of the CG51923-01 gene is detected in cerebellum (CTs=28). In addition, moderate levels of expression of this gene is also seen in number of cancer cell lines derived from tongue, breast, epidermoid carcinoma, lymphoma, bladder, pancreatic, cervical, uterine, and lung cancers. Please see panel 1.1 for further discussion on utility of this gene.

**Panel CNS\_1 Summary:** Ag888 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Please see Panel 1.1 for a



discussion of the potential utility of this gene in treatment of central nervous system disorders.

**Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences**

5 Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can  
10 be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be  
15 silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation,  
20 intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that  
25 resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

30 Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions

identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

#### CG108175-01 SNP data:

Two SNP variants of CG108175-01 were identified and are shown in Table D1.

Table D1. Table data for CG108175-01						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13378176	1541	A	C	494	Gln	Pro
13379410	5329	T	C	0		

#### CG108782-01 SNP data:

One SNP variant of CG108782-01 was identified and is shown in Table D2.

Table D2. Table data for CG108782-01						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified

13379416	511	C	A	133	Pro	Pro
----------	-----	---	---	-----	-----	-----

**CG108801-01 SNP data:**

Six SNP variants of CG108801-01 were identified and are shown in Table D3.

<b>Table D3. Table data for CG108801-01</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379413	556	T	C	124	His	His
13374149	1085	G	A	301	Gly	Arg
13376555	1095	A	G	304	Glu	Gly
13376553	1260	A	G	359	Gln	Arg
13379415	1285	C	T	367	Leu	Leu
13376552	1326	C	T	381	Thr	Ile

5

**CG111815-01 SNP data:**

Two SNP variants of CG111815-01 were identified and are shown in Table D4.

<b>Table D4. Table data for CG111815-01</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379488	1140	C	T	369	Cys	Cys
13379489	1239	C	T	402	Gly	Gly

**CG112813-01 SNP data:**

Three SNP variants of CG112813-01 were identified and are shown in Table D5.

<b>Table D5. Table data for CG112813-01</b>		
<b>Variant</b>	<b>Nucleotides</b>	<b>Amino Acids</b>

	Position	Initial	Modified	Position	Initial	Modified
13379483	92	T	G	29	Phe	Val
13379484	169	G	A	54	Arg	Arg
13378445	1115	G	A	370	Glu	Lys

**CG112881-02 SNP data:**

One SNP variant of CG112881-02 was identified and is shown in Table D6.

<b>Table D6. Table data for CG112881-02</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379428	834	T	C	92	Leu	Leu

5

**CG113377-01 SNP data:**

One SNP variant of CG113377-01 was identified and is shown in Table D7.

<b>Table D7. Table data for CG113377-01</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379431	564	A	G	186	Asn	Ser

**CG123772-01 SNP data:**

One SNP variant of CG123772-01 was identified and is shown in Table D8.

<b>Table D8. Table data for CG123772-01</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
c100.191	1057	A	G	349	Thr	Thr

10

**CG50880-04 SNP data:**

Four SNP variants of CG50880-04 were identified and are shown in Table D9.

<b>Table D9. Table data for CG50880-04</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379422	128	C	T	43	Arg	Trp
13379421	162	T	C	54	Ile	Thr
13379420	423	C	T	141	Ser	Phe
13379419	588	T	C	196	Val	Ala

**CG51923-01 SNP data:**

5 Two SNP variants of CG51923-01 were identified and are shown in Table D10.

<b>Table D10. Table data for CG51923-01</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13379429	2781	C	A	923	Pro	His
13379430	10450	G	A	3479	Pro	Pro

**CG103191-02 SNP data:**

One SNP variant of CG51923-01 was identified and is shown in Table D11.

<b>Table D11. Table data for CG103191-02</b>						
<b>Variant</b>	<b>Nucleotides</b>			<b>Amino Acids</b>		
	Position	Initial	Modified	Position	Initial	Modified
13378180	903	A	G	281	Glu	Glu

10 **CG110725-01 SNP data:**

Two SNP variants of CG110725-01 were identified and are shown in Table D12.

Table D12. Table data for CG110725-01						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
c110.762	344	T	C	94	Asp	Asp
13378565	892	G	A	277	Arg	His

#### CG121519-01 SNP data:

Two SNP variants of CG121519-01 were identified and are shown in Table D13.

Table D13. Table data for CG121519-01						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13379378	228	T	G	76	His	Gln
13379377	1134	C	A	378	Ser	Ser

5

## OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

## CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.
9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
- (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
- b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

- (a) introducing said polypeptide to said agent; and
- (b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
- (b) contacting the cell with a composition comprising a candidate substance; and



- (c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

- (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
- (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
- (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to

a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.
19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61 or a biologically active fragment thereof.
20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61.
21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.
22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61.
23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 61.
24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 61.
25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 61, or a complement of said nucleotide sequence.

26. A vector comprising the nucleic acid molecule of claim 20.
27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.
28. A cell comprising the vector of claim 26.
29. An antibody that immunospecifically binds to the polypeptide of claim 1.
30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.
31. The antibody of claim 29, wherein the antibody is a humanized antibody.
32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:
- (a) providing said sample;
  - (b) introducing said sample to a probe that binds to said nucleic acid molecule; and
  - (c) determining the presence or amount of said probe bound to said nucleic acid molecule,
- thereby determining the presence or amount of the nucleic acid molecule in said sample.
33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.
34. The method of claim 33 wherein the cell or tissue type is cancerous.
35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:
- a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and

- b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 61.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.

REVISED VERSION

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
9 October 2003 (09.10.2003)

PCT

(10) International Publication Number  
**WO 2003/083039 A2**

(51) International Patent Classification <sup>7</sup> : <b>A61K 31/00</b>		US	60/372,141 (CON)
(21) International Application Number:		Filed on	12 April 2002 (12.04.2002)
PCT/US2002/021485		US	60/372,967 (CON)
(22) International Filing Date: 3 July 2002 (03.07.2002)		Filed on	16 April 2002 (16.04.2002)
(25) Filing Language: English		US	60/303,828 (CON)
(26) Publication Language: English		Filed on	9 July 2001 (09.07.2001)
(30) Priority Data:		US	60/361,748 (CON)
60/303,046	5 July 2001 (05.07.2001)	Filed on	5 March 2002 (05.03.2002)
60/303,828	9 July 2001 (09.07.2001)	US	60/304,502 (CON)
60/304,502	11 July 2001 (11.07.2001)	Filed on	11 July 2001 (11.07.2001)
60/305,011	12 July 2001 (12.07.2001)	US	60/324,711 (CON)
60/305,262	13 July 2001 (13.07.2001)	Filed on	25 September 2001 (25.09.2001)
60/305,673	16 July 2001 (16.07.2001)	US	60/361,178 (CON)
60/306,085	17 July 2001 (17.07.2001)	Filed on	1 March 2002 (01.03.2002)
60/307,536	24 July 2001 (24.07.2001)	US	60/358,939 (CON)
60/308,228	27 July 2001 (27.07.2001)	Filed on	22 February 2002 (22.02.2002)
60/308,877	30 July 2001 (30.07.2001)	US	60/322,640 (CON)
60/312,203	14 August 2001 (14.08.2001)	Filed on	17 September 2001 (17.09.2001)
60/322,640	17 September 2001 (17.09.2001)	US	60/363,683 (CON)
60/323,484	19 September 2001 (19.09.2001)	Filed on	12 March 2002 (12.03.2002)
60/323,948	21 September 2001 (21.09.2001)	US	60/305,011 (CON)
60/323,821	21 September 2001 (21.09.2001)	Filed on	12 July 2001 (12.07.2001)
60/324,711	25 September 2001 (25.09.2001)	US	60/305,262 (CON)
60/327,893	9 October 2001 (09.10.2001)	Filed on	13 July 2001 (13.07.2001)
60/331,768	21 November 2001 (21.11.2001)	US	60/373,051 (CON)
60/359,191	21 February 2002 (21.02.2002)	Filed on	16 April 2002 (16.04.2002)
60/358,939	22 February 2002 (22.02.2002)	US	60/373,881 (CON)
60/360,923	28 February 2002 (28.02.2002)	Filed on	19 April 2002 (19.04.2002)
60/361,178	1 March 2002 (01.03.2002)	US	60/305,673 (CON)
60/360,830	1 March 2002 (01.03.2002)	Filed on	16 July 2001 (16.07.2001)
60/361,748	5 March 2002 (05.03.2002)	US	60/307,536 (CON)
60/363,683	12 March 2002 (12.03.2002)	Filed on	24 July 2001 (24.07.2001)
60/363,429	12 March 2002 (12.03.2002)	US	60/373,287 (CON)
60/372,141	12 April 2002 (12.04.2002)	Filed on	17 April 2002 (17.04.2002)
60/372,967	16 April 2002 (16.04.2002)	US	60/323,948 (CON)
60/373,051	16 April 2002 (16.04.2002)	Filed on	21 September 2001 (21.09.2001)
60/373,063	16 April 2002 (16.04.2002)	US	60/359,191 (CON)
60/373,287	17 April 2002 (17.04.2002)	Filed on	21 February 2002 (21.02.2002)
60/373,280	17 April 2002 (17.04.2002)	US	60/306,085 (CON)
60/373,881	19 April 2002 (19.04.2002)	Filed on	17 July 2001 (17.07.2001)
10/187,975	2 July 2002 (02.07.2002)	US	60/323,484 (CON)
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:		Filed on	19 September 2001 (19.09.2001)
US	60/303,046 (CON)	US	60/323,821 (CON)
Filed on	5 July 2001 (05.07.2001)	Filed on	21 September 2001 (21.09.2001)
		US	60/363,429 (CON)
		Filed on	12 March 2002 (12.03.2002)
		US	60/373,280 (CON)
		Filed on	17 April 2002 (17.04.2002)
		US	60/308,228 (CON)
		Filed on	27 July 2001 (27.07.2001)
			[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract:

WO 2003/083039 A2



US 60/308,877 (CON)  
 Filed on 30 July 2001 (30.07.2001)  
 US 60/312,203 (CON)  
 Filed on 14 August 2001 (14.08.2001)  
 US 60/373,063 (CON)  
 Filed on 16 April 2002 (16.04.2002)  
 US 60/360,923 (CON)  
 Filed on 28 February 2002 (28.02.2002)  
 US 60/360,830 (CON)  
 Filed on 1 March 2002 (01.03.2002)  
 US 60/327,893 (CON)  
 Filed on 9 October 2001 (09.10.2001)  
 US 60/331,768 (CON)  
 Filed on 21 November 2001 (21.11.2001)  
 US Not furnished (CON)  
 Filed on 2 July 2002 (02.07.2002)

(71) **Applicant** (for all designated States except US): **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** (for US only): **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apt. 1C, Branford, CT 06405 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US). **CATTERTON, Elina** [FI/US]; 584, Boston Post Road, Madison, CT 06443 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Drive, New Haven, CT 06515 (US). **HJALT, Tord** [SE/US]; 514 Main Street, Apartment 30, East Haven, CT 06512 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **BAUMGARTNER, Jason, C.** [US/US]; 1697 Quinnipiac Avenue, New Haven, CT 06513 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **PEYMAN,**

**John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **SMITHSON, Glenda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **PENA, Carol, E., A.** [US/US]; 604 Orange Street, Number 2, New Haven, CT 06511 (US). **CHAPOVAL, Andrei** [US/US]; 1 Maltby Street, Branford, CT 06405 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US).

(74) **Agent:** **ELRIFI, Ivor, R.;** Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C, One Financial Center, Boston, MA 02111 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority

(48) **Date of publication of this revised version:** 10 June 2004

(15) **Information about Correction:**

see PCT Gazette No. 24/2004 of 10 June 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

# PCT

## DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rule 13ter.1(c) and 39)

Applicant's or agent's file reference 21402-397A	<b>IMPORTANT DECLARATION</b>		Date of mailing (day/month/year)
International application No. PCT/US02/21485	International filing date (day/month/year) 03 July 2002 (03.07.2002)	(Earliest) Priority date (day/month/year) 05 July 2001 (05.07.2001)	
International Patent Classification (IPC) or both national classification and IPC IPC(7): A61K 31/00 and US Cl.: 530/350			
Applicant CURAGEN CORPORATION			

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below.

1. ☐ The subject matter of the international application relates to:
  - a. ☐ scientific theories.
  - b. ☐ mathematical theories
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essential biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☐ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practised on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
 

☐ the description
 ☐ the claims
 ☐ the drawings
3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions prevents a meaningful search from being carried out:
 

☒ the written form has not been furnished or does not comply with the standard.
 ☒ the computer readable form has not been furnished or does not comply with the standard.

4. Further comments:  
Please See Continuation Sheet

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer  Michael Borin Telephone No. (703) 308-0196
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------

Form PCT/ISA/203 (July 1998)

DECLARATION OF NON-ESTABLISHMENT OF  
INTERNATIONAL SEARCH REPORT  
Form PCT/ISA/203 (July 1998)

International application No.  
PCT/US02/21485

4. Further comments:

Since the Sequence Listing was not filed on the international filing date, the sequence listing in written form does not form part of the International application (see PCT Rule 13 ter.1(e)). Regardless of the medium that the sequence listing in written form would have been presented on in response to the invitation, the sequence listing in written form would not under PCT Rule 13 ter.1(e) form a part of International application.